Metabolomics-based method development for biomarker identification

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades

einer Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von **Ivana Blaženović**
aus Zagreb / Kroatien

1. Referent: Professor Dr. Dieter Jahn
2. Referent: Professor Dr. Karsten Hiller

eingereicht am: 24.10.2016
mündliche Prüfung (Disputation) am: 10.01.2017

Druckjahr 2017

Vorveröffentlichungen der Arbeit

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Ivana Blaženović, Slobodan Obrenović, Daniel Stoessel, Louisa Roselious, Violetta Mass, Nicolas Schauer, Peter Hammerer, Rebekka Biedendieck, Dieter Jahn, Martina Jahn. Prognostic model of urinary tract infections (submitted, 2016).

Ivana Blaženović, Tobias Kind, Hrvoje Torbašinović, Slobodan Obrenović, Sajjan Mehta, Hiroshi Tsugawa, Tobias Wermuth, Nicolas Schauer, Dieter Jahn, Oliver Fiehn. Comprehensive re-analysis of CASMI 2016 results: *in silico* MS/MS fragmentation tools must be combined with database boosting to achieve 92% accuracy (submitted, 2016).

Ivana Blaženović, Jelle Zwaag, Daniel Stoessel, Josephine M Worseck, Nicolas Schauer, Dieter Jahn, Peter Pickkers, Matthijs Kox. The metabolic response to voluntary activation of the sympathetic nervous system and attenuation of the innate immune response in humans (in preparation).

Patent

Patent for diagnosis of urogenital tract infections

Inventors: Dieter Jahn, Nicolas Schauer, **Ivana Blaženović**

Ref: 856-3 EP (2016)

Tagungsbeiträge

Blaženović I., Burghartz M., Schauer N., Jahn D.: Bacterial Metabolites for a New Cost-Effective Method of Diagnosing Urinary Tract Infections (poster), Microbiology and Infection - 4th Joint Congress of DGHM and VAAM, Dresden, Germany (October, 2014).

Ivana Blaženović, Slobodan Obrenović, Daniel Stoessel, Nicolas Schauer, Peter Hammerer, Martina Jahn, Rebekka Biedendieck, Dieter Jahn.: Prognostic model of urinary tract infections (oral presentation), Annual Conference of the Association for General and Applied Microbiology (VAAM), Jena, Germany (March, 2016).

Ivana Blaženović, Slobodan Obrenović, Daniel Stoessel, Nicolas Schauer, Peter Hammerer, Martina Jahn, Rebekka Biedendieck, Dieter Jahn.: Prognostic model of urinary tract infections and structure elucidation (oral presentation), West Coast Metabolomics Center Meeting, UC Davis Genome Center – Metabolomics, Davis, CA, U.S.A. (April, 2016).

Ivana Blaženović, Tobias Kind, Hrvoje Torbašinović, Slobodan Obrenović, Sajjan Mehta, Hiroshi Tsugawa, Tobias Wermuth, Nicolas Schauer, Dieter Jahn, Oliver Fiehn: Compound identification in metabolomics: *in silico* fragmentation algorithms (oral presentation), West Coast Metabolomics Center Meeting, UC Davis Genome Center – Metabolomics, Davis, CA, U.S.A. (August, 2016).

For my family

Table of Contents

METABOLOMICS-BASED METHOD DEVELOPMENT FOR BIOMARKER IDENTIFICATION ..... - 1 -

TABLE OF CONTENTS.....II

II ABBREVIATIONS VI

1 INTRODUCTION..... - 9 -

1.1. Metabolomics..... - 9 -

1.2. Untargeted versus targeted metabolomics..... - 11 -

1.3. Limitations and challenges of the metabolomics approach ..... - 13 -

1.4. Microbiology and metabolomics..... - 14 -

1.5. Identification of the microbes ..... - 15 -

1.6. Urinary tract infections ..... - 17 -

1.7. Aim of the study ..... - 20 -

2 PROGNOSTIC MODEL OF URINARY TRACT INFECTIONS ..... - 21 -

2.1. Abstract Part I..... - 22 -

2.2. Introduction – Part I..... - 23 -

2.3. Materials and Methods – Part I..... - 26 -

2.3.1. Human urine samples - 26 -

2.3.2. 16S rRNA sequence analysis for identification of the uropathogens using single strand conformation polymorphism (SSCP) analysis - 26 -

2.3.3. Sample preparation for LC-MS analysis - 27 -

2.3.4. Metabolomics measurements - 28 -

2.4. Results and Discussion – Part I ..... - 30 -

2.4.1. Identification of the bacteria in urine samples - 30 -

2.4.2. Urinary metabolomics - 31 -

2.4.3. Validation of putative biomarkers - 34 -

2.5. Conclusion – Part I ..... - 40 -

2.6. Acknowledgement – Part I ..... - 41 -

3 COMPREHENSIVE RE-ANALYSIS OF CASMI 2016 RESULTS: *IN SILICO* MS/MS FRAGMENTATION TOOLS MUST BE COMBINED WITH DATABASE BOOSTING TO ACHIEVE 92% ACCURACY – PART II ..... - 42 -

3.1. Abstract Part II ..... - 43 -

3.2. Introduction – Part II ..... - 44 -

3.3. Materials and Methods – Part II ..... - 46 -

3.3.1. Tandem mass spectral input data - 46 -

3.3.2. Query compounds from ChemSpider - 46 -

3.3.3. Software settings - 47 -

3.3.3.1. MS-Finder - 47 -

3.3.3.2. CFM-ID - 47 -

3.3.3.3. MetFragCL - 48 -

3.3.3.4. MAGMa+ - 48 -

3.4. Results – Part II ..... - 50 -

3.4.1. CASMI Category 2 (Best Automatic Structural Identification - *In Silico* Fragmentation Only): parameter optimization and development of voting/consensus model - 50 -

3.4.1.1. *In silico* performance using the training set - 50 -

3.4.1.2. Voting/consensus model - 51 -

3.4.1.3. Voting/consensus model applied to *in silico* results - 53 -

3.4.2. CASMI Category 3 (Best Automatic Structural Identification - Full Information): application of voting/consensus model - 53 -

3.4.2.1. Compound database search and importance ranking - 53 -
3.4.2.2. MS/MS search settings - 54 -
3.4.2.3. Sorting according to importance using ChemSpider ID - 54 -
3.4.2.4. Final scoring model for the validation set - 55 -
3.4.2.5. Validation set performance for Categories 2 and 3 - 58 -

3.5. Discussion – Part II ..... - 60 -
3.5.1. Training set performance - 60 -
3.5.2. Validation set performance and structure diversity - 61 -
3.5.3. Voting/consensus model performance - 62 -
3.5.4. MS/MS data quality - 62 -
3.5.5. Software use and performance - 63 -
3.5.6. Continuous integration and design of experiment - 63 -

3.6. Conclusion – Part II ..... - 65 -

3.7. Acknowledgement – Part II ..... - 66 -

3.8. Supplemental material – Part II ..... - 67 -
3.8.1. Methods - 68 -
3.8.2. Software settings ..... - 68 -
3.8.2.1. MS-Finder ..... - 68 -
3.8.2.2. MetFragCL ..... - 70 -
3.8.2.3. MAGMa+: ..... - 70 -
3.8.2.4. CFM-ID ..... - 71 -
3.8.2.5. MS/MS database search ..... - 71 -
3.8.2.6. CSI: FingerID ..... - 73 -

4 THE METABOLIC RESPONSE TO VOLUNTARY ACTIVATION OF THE SYMPATHETIC NERVOUS SYSTEM AND ATTENUATION OF THE INNATE IMMUNE RESPONSE IN HUMANS ..... - 76 -

4. 1. Introduction – Part III ..... - 77 -

4.2. Materials and methods – Part III ..... - 78 -

4.2.1. Subjects and experimental design - 78 -
4.2.2. Sample preparation for LC-MS analysis - 79 -
4.2.3. LC-MS analysis - 79 -
4.2.4. Cytokine and catecholamine determination - 80 -
4.2.5. Raw data processing and data analysis - 80 -

4.3. Results – Part III ..... - 82 -

4.4. Discussion and Conclusion – Part III ..... - 91 -

4.5. Acknowledgement – Part III ..... - 93 -

5 PATENT FOR DIAGNOSIS OF UROGENITAL TRACT INFECTIONS: PART IV ..... - 94 -

5.1. Summary of the invention ..... - 94 -

5.2. Acknowledgement – Part IV ..... - 96 -

6 SUMMARY ..... - 97 -

7 ZUSAMMENFASSUNG ..... - 98 -

8 OUTLOOK ..... - 100 -

9 REFERENCES ..... - 101 -

APPENDICES ..... - 112 -

ACKNOWLEDGMENT ..... - 143 -

CURRICULUM VITAE ..... - 147 -

II Abbreviations

°C	Celsius (°C)
ADP	adenosine diphosphate
ATP	adenosine triphosphate
AUC	area under the curve
bp	base pair
CASMI	critical assessment of small molecule identification
CFM	competitive fragmentation modelling
CID	collision induced dissociation
CL	command line
Da	Dalton
db	database
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
e.g.	*exempli gratia* (for instance)
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionization
et al.	*et alia* (and others)
FA	formic acid
g	→ *centrifugation*: earth gravity, → *weight*: gram
G	giga
GC	gas chromatography
h	hour
H_2O_{dest}	*aqua destillata*

HILIC	hydrophilic interaction liquid chromatography
HMDB	Human Metabolome Database
Hz	hertz
InChI	International Chemical Identifier
IOKR	Input Output Kernel Regression
k	kilo
kDa	kilo Dalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
L	liter
LB	Lysogenic broth
m	milli
µ	micro
M	molar (mol/l)
METLIN	Metabolomics database
min	minute
M_r	relative molecular mass
MS	mass spectrometry
MS/MS	tandem mass spectra
MSC	Molecular Structure Correlator
n	nano
NADH	nicotinamide adenine dinucleotide, reduced form
NMR	nucleic magnetic resonance
Pa	Pascal
PCA	principal component analysis
PCR	polymerase chain reaction
Ppm	parts per million

rdp	raw data processing
ROC	receiver operating characteristics
rpm	rotations per minute
rRNA	ribosomal ribonucleic acid
RT	Retention time
s	second
SMILES	simplified molecular-input line-entry system
SSCP	single strand conformation polymorphism
T	temperature
TCA	tricarboxylic acid cycle
U	unit
UMDB	Urine Metabolome Database
UPLC	ultra pressure liquid chromatography
UTI	urinary tract infection
UV	ultraviolet
V	volt
v/v	volume per volume
vol.	volume
w/v	weight per volume

1 Introduction

1.1. Metabolomics

Metabolomics is defined as the study of the metabolite composition of a living cell, a biological fluid or a tissue by the means of mass spectrometry. Metabolome, the analysis of the complete set of metabolites, is one of the fastest developing disciplines of current life sciences. To date, metabolomics is considered as a major tool in addressing challenging research objectives including precision and personalized medicine. Metabolomics constitutes the downstream products of genomic, epigenomic, transcriptomic, and proteomic processes as shown in Figure 1.

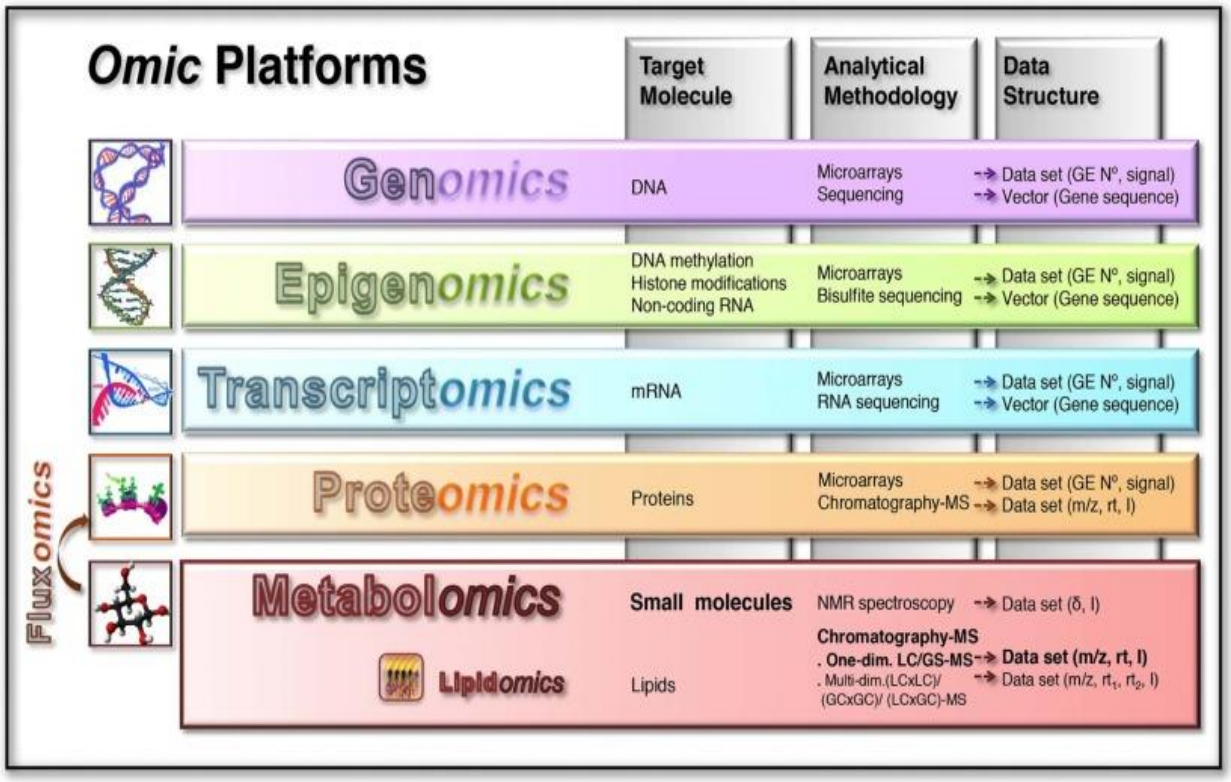


Figure 1: Central dogma of the "omics" studies (adapted from Gorrochategui *et al*., 2016).

The end products of enzymatic reactions or their intermediates are defined as metabolites. They represent the most informative snapshot of the biochemical activity of an organism [1]. The present technologies in metabolomics are allowing the studies of complex

biological samples and the detection of thousands of metabolites [2]. Application of metabolomics is broad and it covers research areas like nutrition [3], drug discovery [4], plant biology [5] and a wide range of the studies of human diseases [6]. One of the most active areas of development in metabolomics is the biomedical field. The use of metabolomics is supporting development of prognostic and diagnostic biomarkers that aid medical practitioners to achieve a level of information about biological systems in humans required for safe diagnosis. Consequently, this experimental approach opens up new perspectives in many fields of medicine [7]. To date, metabolomics data is generated mostly by two techniques: nuclear magnetic resonance (NMR) and mass spectrometry (MS) [8]. However, MS is by far the dominating technology in metabolomics studies. It is described as an analytical technique that is recording spectral data in the form of relative intensity and mass-to-charge ratio (m/z) for compounds of interest. First, metabolites of interest need to be extracted from the investigated cells or tissues. Second, biological extracts need to be ionized for the mass spectrometer to generate signals for the analyzed metabolites. A wide range of ionization techniques for mass spectrometry are currently available. Prior to ionization of the biological samples usually there is a separation step in metabolomics. Fractionation of the biological samples reduces its complexity. Most commonly used techniques for this purpose are liquid and gas chromatography columns (LC and GC), respectively [9].

Chromatography is a separation technique in which a mobile phase contains compounds to be separated is moved over chemically defined stationary phase. Separation is based on the differential interaction between the stationary phase and the different metabolites in the mobile phase. In liquid chromatography (LC) mobile phase is a liquid, in gas chromatography (GC) is gaseous. Differences in rates of movement along the stationary phase are caused by the different chemical properties of the metabolites. Retention time represents the time needed for each metabolite to pass along the stationary phase, and deduced m/z MS values are used to generate the two axes of the GC-MS and LC-MS spectral data [10].

1.2. Untargeted versus targeted metabolomics

Metabolomics analysis can be divided into untargeted and targeted approaches, depending on the design of the experiments [11]. Untargeted metabolomics focuses on the analysis of all the detectable metabolites in a sample. A flow chart of a typical LC-MS metabolomics experiment is shown in Figure 2. Untargeted and targeted approaches are visualized.

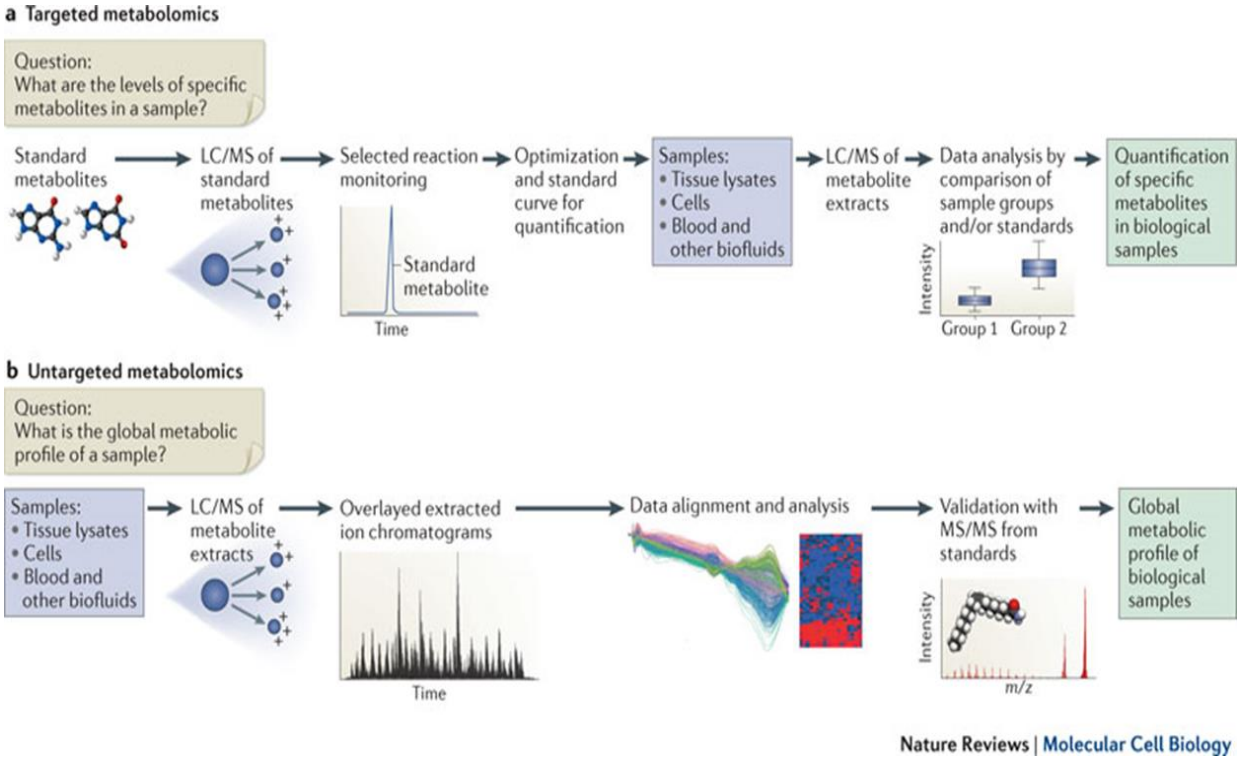


Figure 2: Analysis workflow in untargeted metabolomic studies. This figure shows the different steps of the metabolomic analysis pipeline (adapted from Patt *et al.*, 2012).

The workflow consists of the experimental design, the extraction protocol, data acquisition, processing and analysis, and metabolite identification to finally allow biological interpretation. The aim of the investigation determines the experimental design and data acquisition methods [12]. These studies are characterized by generating large amounts of data of high complexity which require high performance bioinformatics tools to process the data [13, 14]. Data processing was improved enormously over the years due to the development of automated procedures such as MetAlign (http://www.metalign.wur.nl), MZmine (http://mzmine.sourceforge.net/) and XCMS (http://metlin.scripps.edu/download/). These programs are designed to extract relevant

information automatically from batches of crude chromatographic data. They allow for the rapid processing of thousands of data points, background noise subtraction, signal recognition and alignment. This processing is needed for the transformation of the untargeted metabolomics data into straightforward metabolite identifications. Output of these tools is usually a data table listing the amount of each signal in all samples [15].

Due to its comprehensive nature, untargeted metabolomics must be connected to advanced chemometric techniques in order to reduce extensive datasets into a smaller set of manageable signals, such as multivariate analysis. These signals then require annotation using either *in silico* libraries or experimental investigation with identification step by means of analytical chemistry. Untargeted analysis offers the opportunity for novel target discovery including the discovery of unknown chemical compounds, as coverage of the metabolome is only restricted by the methodologies of sample preparation and the sensitivity and specificity of analytical instrumentation [16-18]. By contrast, targeted metabolomics is the measurement of defined groups of metabolites. Targeted metabolomics is of high importance, particularly in systems medicine, systems biology and biomarker validation where metabolite quantification is of importance. These methods target a limited number of known metabolites, often relying on triple quadrupole mass spectrometry. Data processing for targeted analysis is moreover simplified as the user already knows which metabolite is to be expected. In addition, it is possible to set up a specific processing method for the assay identifying metabolites according to retention time and/or fragmentation pattern. Such approaches have been developed for instance to rapidly quantify eleven key gut microbiota-host co-metabolites within only 10 minutes [19]. Similarly, the development and validation of an assay for the quantification of ten metabolites associated with the Krebs cycle, purine degradation, and oxidative stress in human urine was achieved [20]. Chromatography can be used to ensure the separation of isobaric species (for example, the metabolites citrate and isocitrate) or species that fragment in such a way that they resemble other species (for example ATP may fragment under certain conditions to resemble ADP or AMP in terms of the ions produced) to further enhance the robustness and reliability of the method [21].

1.3. Limitations and challenges of the metabolomics approach

The first major challenge in metabolomics is the identification of the measured metabolites. In mass spectrometry-based studies metabolite spectral databases are essential for metabolite identification. The most common metabolite identification approach is done by querying metabolomics databases for the neutral or ionized molecular mass values of the identified peaks using a tolerance window [22]. The quality of the stored data in these databases is critical for the performance of identification algorithms [23]. Assuming no prior knowledge of the measured metabolites as in untargeted analysis, each peak m/z value can lead to multiple plausible neutral molecular masses that can represent different ionization adducts (H+, Na+, K+...). This multiplicity often results in a high number of false positive identifications [24, 25]. Second major challenge in metabolomics studies is the ion suppression. Ion suppression appears as one particular manifestation of matrix effects, which are influencing the extent of analyte ionization. This change often is observed as a loss in response, thus, the term ionization suppression. However, it also can be observed as an increase in the response of the desired analyte depending upon the type of sample [26]. Furthermore, subtle variations between tested individuals in biomedical studies can result in large variations of the observed metabolite concentrations. The corresponding flux of the metabolites is influenced by genetic and environmental conditions. Some of the environmental factors include diet, stress, xenobiotic use, lifestyle, and disease. Genetic variation can result from differences in gender, epigenetics, and gene polymorphisms. All this has to be taken into consideration when designing a metabolomics experiment [27].

The third major challenge lies in the decision which instrumentation is employed to analyze the samples. It is known that no single technical approach is able to reveal the entire metabolome. This is due to the heterogeneity of the chemical constitution of the metabolites and the specific limitations of each instrument [28, 29].

By far the most challenging part of a metabolomic study is the confirmation of discovered potential biomarkers. This is essential towards the understanding of biological changes occurring within the system and is until now considered a major bottleneck in metabolomics investigations. Technological advances with better databases could help solve this problem. In fact, metabolite identification from LC-MS data is more

problematic as the databases are much less developed than those for GC-MS analysis. These include the Human Metabolome Database (HMDB) (http://www.hmdb.ca) and the METLIN Metabolite Database (http://metlin.scripps.edu). It is also difficult to compare data sets between laboratories due to different used analytical instrumentation and experimental setup [30, 31].

1.4. Microbiology and metabolomics

Infectious diseases continue to be a major health concern worldwide: hepatitis C, AIDS, malaria, tuberculosis (TB) and urinary tract infections are ongoing pandemics [32]. However, microorganisms are also of central importance for the ecological process keeping our world in balance. They are also major tools of biotechnology. For all this processes metabolomics can provide valuable information about the molecular processes involved. Metabolomics alone, however, is inadequate to understand the cellular metabolic activity: flux measurement, genomic, proteomic and transcriptomic data must be added to obtain a systems biology view of the biological process of interest. Clearly, biological processes require an integrated approach to study the sum of all systems [33, 34]. The new omics technologies, computational methods and genome information are needed for the successful implementation of systems biology. Technologies such as the next-generation sequencing have opened the door to elucidate total transcriptomes, encompassing microRNAs, long noncoding RNAs and mRNAs [35]. Microorganisms are ideal for conducting systems biology studies because they are relatively easy to operate and manipulate. Thus, microbial metabolomics integrates biological information into systems microbiology to aid the understanding of microbial interactions and cellular functions [36]. The influence of the microbes on human health has been left unstudied leaving human development, physiology, immunity and nutrition almost entirely unknown.

1.5. Identification of the microbes

Traditionally, bacterial identification in microbiology laboratories was performed using phenotypic characterization, including Gram staining in combination with a variety of biochemical tests, taking into account culture requirements and growth characteristics.

However, these methods of bacterial identification are known to have major limitations. First, often it is observed that there are organisms with biochemical characteristics that do not fit into the patterns of any known genus and species. Second, traditional methods cannot be applied to uncultivable organisms which represent over 95% of all known bacteria. Third, identification of anaerobes and mycobacteria require additional expertise and equipment, which are not available in most clinical microbiology laboratories [37]. However, these problems can be avoided using a technology, known as 16S rDNA sequencing. The microbiologist George Fox and colleagues described their technology in 1980 that revolutionized the way prokaryotic organisms are characterized, classified, and identified. In their study, molecular techniques were used to determine the nucleotide sequence of ribosomal RNA from various prokaryotic organisms in order to assess their relative position in the evolutionary tree [38].

Since that time, such analyses have become commonplace in microbiology research laboratories as a tool for classification of microbial organisms. According to 16S rRNA gene analyzes, life on earth has now been classified into three domains: Bacteria, Archaea, and Eucarya. Each domain has been subdivided into at least two kingdoms by further genotypic and/or phenotypic analyses [39, 40]. The relationship between the major branches of life, the Archaea, Bacteria and Eucarya, as well as the major branches within these domains is based on determined 16S rRNA/18S rDNA gene sequences, as shown in Figure 3.

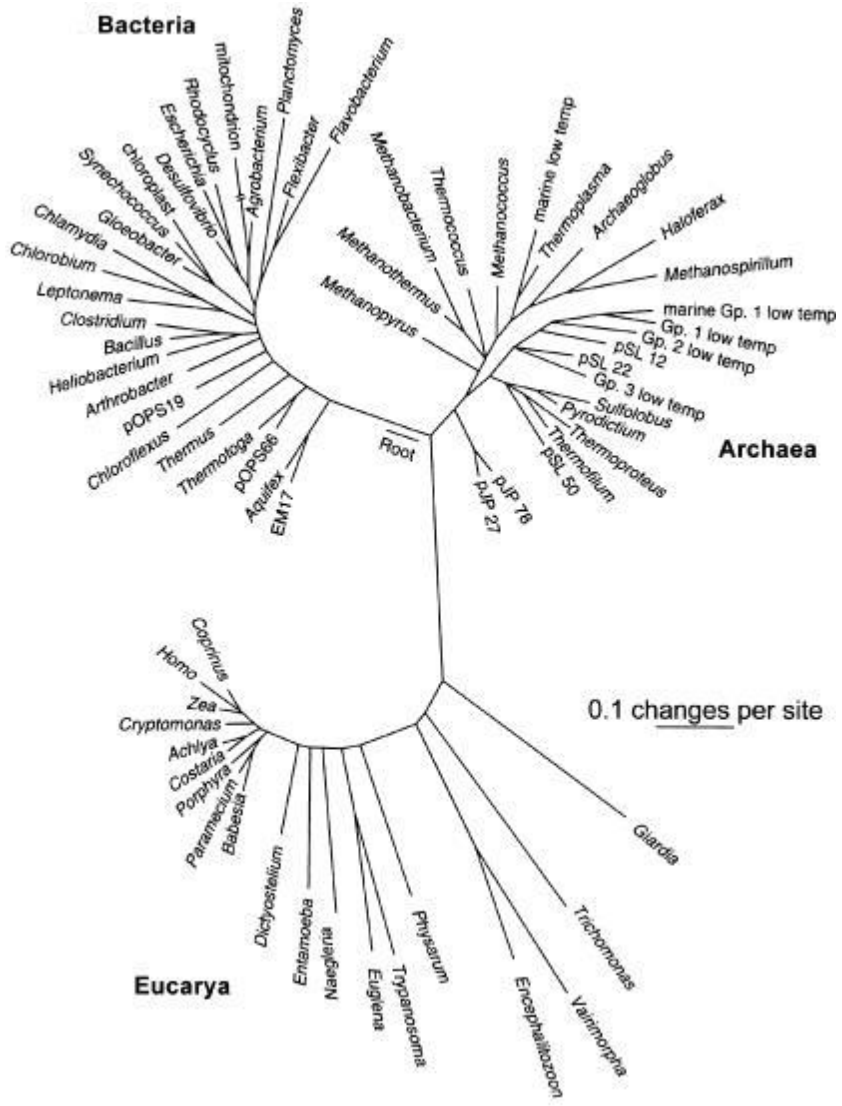


Figure 3: Universal phylogenetic tree based on the 16S rRNA gene sequence comparisons (adapted from Woese *et al.*, 1985).

This approach can also be used to determine the structure of microbial communities [38, 41]. Among the three domains of life, Bacteria, Archaea and Eucarya, the largest amount of rRNA gene sequencing work is related to bacteria. Using obtained 16S rRNA gene sequences, many bacterial genera and species had to be renamed and reclassified. Most importantly, this technique made it possible to classify uncultivatable bacteria [42]. Due to the increasing availability of PCR and DNA sequencing facilities, 16S rRNA gene sequencing is not only being used in research but it plays a major role in clinical microbiology laboratories worldwide.

1.6. Urinary tract infections

Urinary tract infections (UTIs) are considered some of the most common bacterial infections, affecting 150 million people each year worldwide [43]. UTIs are clinically classified as uncomplicated and complicated. Uncomplicated UTIs are characterized by no structural or neurological urinary tract abnormalities and they affect individuals that are otherwise healthy [44]. The UTIs are described as complicated if they are associated with factors that compromise the urinary tract or host defense such as urinary obstruction, urinary retention caused by neurological disease, immunosuppression, renal failure or transplantation and the presence of foreign devices as calculi, indwelling catheters or materials used for drainage [45, 46]. Bacteria in UTIs are both Gram-negative and Gram-positive bacteria and certain fungi, as shown in Figure 4. Uropathogenic *Escherichia coli* (UPEC) is found to be most common causative agent for both uncomplicated and complicated UTIs. Other agents associated with UTIs are *Klebsiella pneumoniae*, *Staphylococcus aprophyticus*, *Enterococcus faecalis*, *Streptococcus* spp, *Proteus mirabilis*, *Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Candida* spp [47, 48]. The challenge in diagnosis and treatment of UTI lies in the inconsistent nature and vagueness of the presenting illness. UTIs are commonly treated with broad-spectrum antibiotics which can lead to long-term alteration of the normal microbiota of the vaginal and gastrointestinal tract and in the development of multidrug-resistant microorganisms.

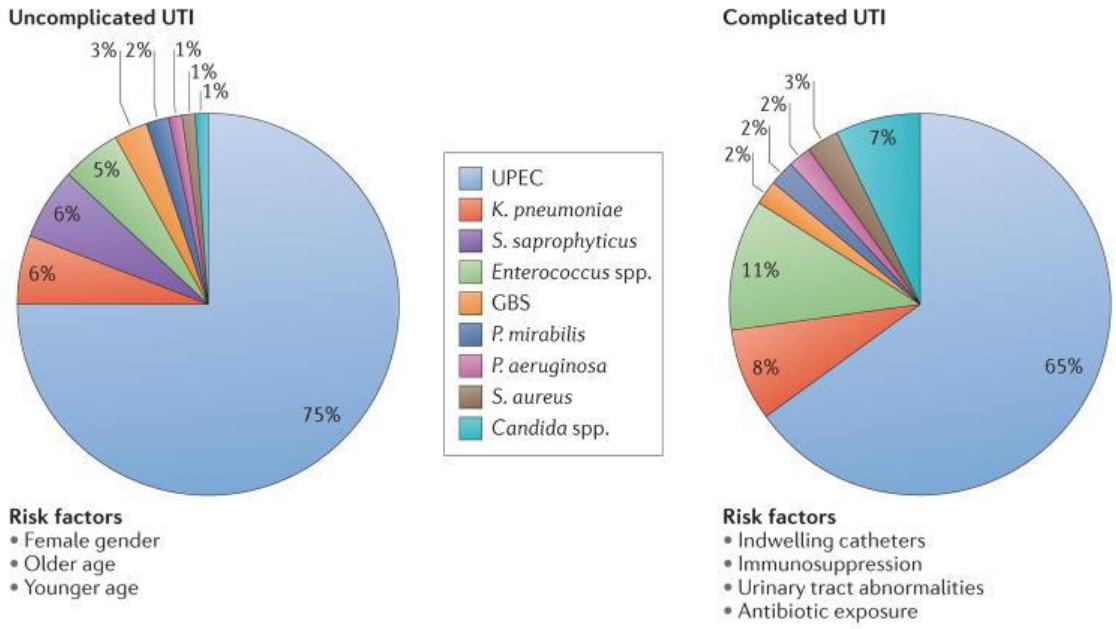


Figure 4: Epidemiology of urinary tract infections (adapted from Ana L. Flores-Mireles *et al*, 2015).

Moreover, high rates of recurrent UTIs suggested that antibiotics are not an effective therapy for all UTIs [49]. Certain clinical practice guidelines define that UTI as the presence of >50,000 colony forming units/mL of a single uropathogen when analyzed by culturing the corresponding urine [50]. Knowing that identification is often not possible by phenotypic methods. It is more precise by 16S rRNA gene sequencing. 16S rRNA gene sequencing has become the reference method for bacterial taxonomy and identification and clinicians need the bacteria identity in order for accurate diagnosis and optimal antibiotic therapy [51]. However, these methods are time-consuming and expensive. A test that could aid clinicians in fast accurate and inexpensive microbial determination of the diagnosis to judge the severity of UTIs would be of great benefit. Metabolomics can be used to quantify specific chemical constituents within a body fluid [52]. Metabolomics is an attractive technology because of its ability to non-invasively provide both qualitative and quantitative measurements, while simultaneously studying thousands of compounds in the same biologic fluid. Urine is an excellent biological fluid for various medical studies because of its easy availability and rich chemical composition with hundreds of metabolites already identified and quantified by LC-MS [53, 54]. Some of the metabolites that are indicative of UTIs that were reported in recent metabolomics studies are acetic acid/creatinine, 6-hydroxynicotinic acid, acetate, lactate, succinate,

formate, glucose, fructose and alanine levels. Those levels were measured in cohort studies and the infected urine samples were usually infected by a single pathogen [55-58]. However, the levels of these general and often found metabolites are of subject to other environmental factors which questions their efficient utilization as specific biomarkers for UTIs. Therefore, we aimed at the development of a metabolomics approach for the identification of highly specific biomarkers safely indicating UTIs.

1.7. Aim of the study

This thesis aimed at employing metabolomics approaches at three research topics:

i. The identification of highly specific biomarkers in human urine which safely indicate urinary tract infections (UTIs). For this purpose, a metabolomics approach was planned where the metabolome of the urine of UTI patients was compared with those of healthy persons. For the safe and efficient identification of urine metabolites indicating UTI the mathematics of the bioinformatics-based identification strategy using the obtained LC-MS data had to be systematically optimized. Obtained potential biomarkers were to be discussed in the context of the current knowledge.

ii. Performance evaluation of four publicly available *in silico* fragmentation tools that aid structure elucidation in metabolomics and that participated in the 2016 Critical Assessment of Small Molecule Identification (CASMI) challenge. Furthermore, development of voting/consensus model to improve the outcomes of each of the existing tools (MetFragCL, CFM-ID, MAGMa+ and MS-Finder).

iii. Metabolite profiling of 12 healthy individuals exposed to endotoxin LPS over the period of 8 hours. In addition, metabolomic profiling of 12 healthy individuals that underwent “iceman training” and were exposed to endotoxin LPS. Metabolic response was investigated between the two groups.

2 Prognostic model of urinary tract infections

Ivana Blaženović [†, ‡], Slobodan Obrenović [§], Daniel Stoessel [‡], Nicolas Schauer [‡],

*Louisa Roselius [†J], Violetta Mass [†], Peter Hammerer [¥], Rebekka Biedendieck [†J], Dieter Jahn [†,J], Martina Jahn †**

[†] Institute of Microbiology, Technische Universität Braunschweig, Germany

[‡] Metabolomic Discoveries GmbH, Potsdam, Germany

[§] Inovatus Ltd, Zagreb, Croatia

[¥]Braunschweig Municipal Hospital GmbH, Urology and Uro-oncology Clinic, Braunschweig, Germany

[J]Braunschweig Integrated Centre of Systems Biology (BRICS), Technische Universität Braunschweig

*Corresponding author

Dr. Martina Jahn

m.jahn@tu-bs.de

My contribution: design of experiments (with D.J., M.J. and N.S.); experimental procedure of metabolomics and SSCP analyses; raw data processing and statistical analysis; first draft of the manuscript.

2.1. Abstract – Part I

The current methodology for urinary tract infection (UTI) diagnosis includes cost-intensive patient symptomology, anamnesis and classical culture-based microbiological analysis. Clinical parameters that lead to uncertainty are asymptomatic bacteriuria, sample contamination and non-infectious inflammatory conditions. In order to develop simple, direct and reliable diagnostic tools, a metabolomics approach using urine samples from 92 patients was carried out. Single Strand Conformation Polymorphism (SSCP) analyses were performed to identify the organisms causing the UTI. Data obtained from comprehensive LC-MS analyses of the training set (28 healthy and 28 infected individuals) were used to establish a reliable prognostic model. Using metabolomics data from 20 healthy and 16 infected individuals as validation set, three putative biomarkers, thieno-(2, 3-c) pyridine, heneicosanoyl-glycero-3-phospho-(1'-glycerol) and metabolite with molecular formula $C_9H_{17}NO$ of the training set were identified as statistically significant to represent the specific host response to the infection. No obvious relation of these metabolites to the identity of the infection-causing bacteria was discovered. Obtained results were confirmed in the validation set showing not only the reproducibility of the analytical platform but also the biological significance of the detected metabolites. The potential benefit of these metabolites as putative diagnostic biomarkers was validated through additional analysis, combinatorial receiver operating characteristic (ROC) curve and could be useful in novel diagnostic strategies.

Keywords

Urinary tract infection, metabolomics, *in silico* fragmentation, mass spectrometry, prognostic model, linear regression, putative biomarker

2.2. Introduction – Part I

Urinary tract infections (UTIs) are among the most common diseases worldwide. An estimated 150 million cases of UTIs occur globally and account for medical expenditures of approximately 1000 €/patient making it one of the biggest health concerns of today [59]. For women, the lifetime risk of developing a UTI is greater than 50%, while only 20% of all men are affected [60]. Additionally, UTIs are generally self-limiting, but have a propensity to recur in individuals without anatomical or functional abnormalities [61, 62]. UTIs are generally treated with antibiotics and are consequently prone to the selection for antibiotic resistant uropathogens and commensal bacteria.

Moreover, antibiotic treatment adversely affects the gut and vaginal microbiota [63]. Uropathogens produce adhesins, siderophores and toxins which enable them to invade and colonize the urinary tract [64, 65]. *Escherichia coli* still remains the most common infecting organism in patients with uncomplicated UTI [66] although other pathogens are present including *Aerococci* [67], *Proteus*, *Staphylococci*, *Pseudomonas*, *Enterococci* and *Klebsiella* [68]. Cultivation of bacteria from urine samples combined with appropriate observation of patients' symptoms are currently the gold standard for diagnosis. However, it is not uncommon to treat UTI solely based on the observed symptoms with broad spectrum antibiotics [69]. Non-cultivable bacteria, contamination of collected urine samples and misinterpretation of asymptomatic UTI remain the major drawbacks in current clinical diagnosis [70]. This often results in patients' unnecessary antibiotic exposure and overtreatment. Consequently, novel methods and approaches for fast, reliable and inexpensive UTI diagnosis are much needed [71, 72]. The interplay between uropathogens and their hosts has been studied for decades using many different experimental approaches [73-75]. Although much has been learned from such studies, most have focused on the global effects of infection, often strain specific ones and not on the discovery of potentially reliable biomarkers which are validated in another sample set. A recent study described an NMR-based urinalysis using a quantitative metabolomics approach to evaluate the microbial-human co-metabolite trimethylamine for its function as a diagnostic biomarker for the detection of UTI's in patients [76].

The application of mass spectrometry in clinical environments becomes more common due to its high resolution analytical power. However, efficient and reliable utilization of

mass spectrometry in clinical routine application requires sophisticated data interpretation [77, 78]. Current methods include support vector machines [79], genetic algorithms [80] and other machine learning algorithms [81, 82]. Predictive modelling for an outcome with multiple variables are usually employed to automate the covariate selections from high dimensional mass spectrometry data (HDOD) [83, 84].

Therefore, hybrid approaches that integrate data pattern discovery and regression analysis in order to retain desired features of both analytical approaches are utilized. The development of a prognostic model usually involves two major steps: model building and validation. The first goal is to identify a group of "exemplars" representative of the data's subjects' pattern. These are typically gained through clustering analysis via unsupervised learning [85, 86]. The obtained clustering pattern identifies informative exemplars which provide valuable associations for an outcome using a subsequent "supervised learning" analysis. In this study, the prognostic model is built via linear regression with the method of least squares based on a training set. For every metabolite in this set a p-value is computed, which indicates the significance of difference between infected and non-infected individual metabolites. After model building the significant metabolites are validated using a validation set.

The quality of a prognostic model is evaluated through its performance, i.e. the ability of the model to correctly predict the prognosis of a patient based on the observed predictors. A few examples of prognostic models include: The Nottingham Prognostic Index, which is used to determine prognosis following breast cancer surgery [87] and validated biomarkers used for the prediction of future cardiovascular events [88]. Receiver operating characteristic (ROC) curve analysis is a standard method for assessing the performance of the prognostic models. It is widely accepted as a statistically valid and objective means for determining the clinical utility of a biomarker in metabolomics studies. The area under the curve (AUC) of the ROC curve facilitates the identification of sensitive and specific biological markers [89].

The aim of this study was to identify potential biomarkers in urine samples of UTI patient's indicative of an infection event by using metabolomics as a platform regardless of the pathogen type. For this purpose, we created a prognostic model with urine samples from healthy and UTI patients. Data from the urine samples were used in a training set

and another unlabeled, randomized set of samples was used for validation purposes. The randomized validation set was utilized to avoid batch effects, model overfitting and to examine the significance of the results, as reproducibility still remains a bottleneck. Metabolomics research struggles with low identification rates compared to proteomics [90]. One of the reasons is that in many cases reference tandem mass spectra are not available. This fact opened the window for a very vibrant development of *in silico* fragmentation algorithms that aid structure elucidation [91, 92].

In this paper we describe how *in silico* fragmentation tool, Molecular Structure Correlator (MSC) (Agilent, USA), aids annotation of MS/MS spectra of the three statistically significant compounds when the reference spectra is missing, by predicting the fragmentation pattern of the molecule, comparing it with the experimental MS/MS spectra and querying the publicly available databases. General workflow of this study is described in the Figure 5.

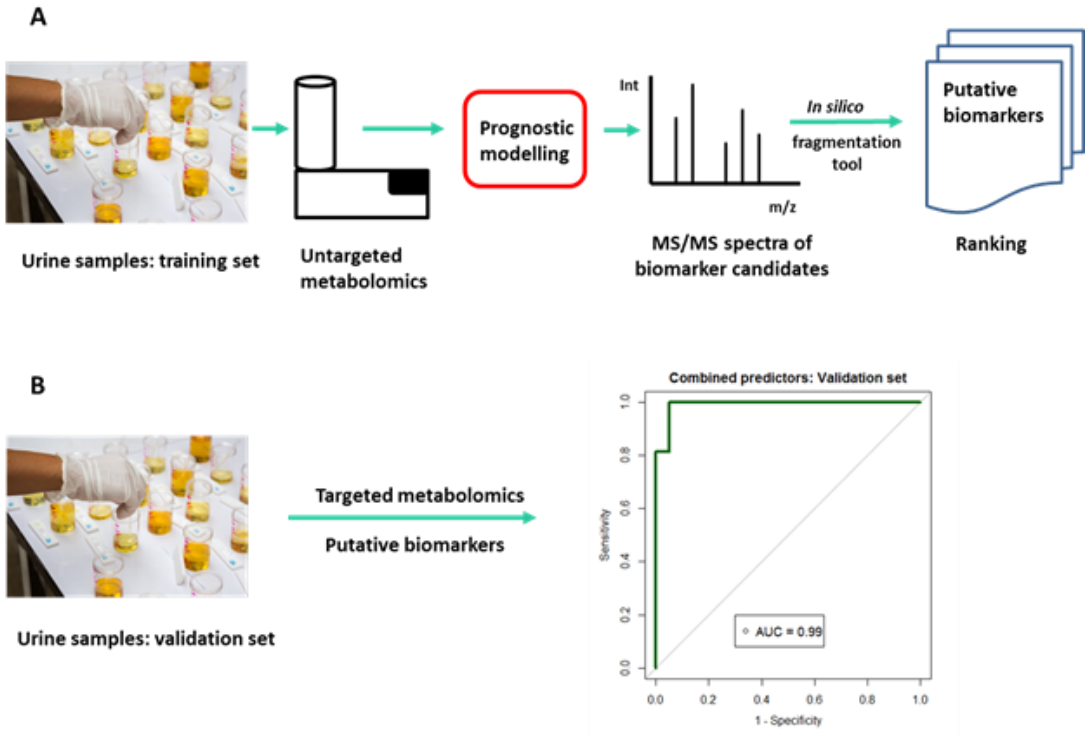


Figure 5: General workflow

2.3. Materials and Methods – Part I

2.3.1. Human urine samples

Urine samples were collected from February to May 2015 in a Urology clinic in Braunschweig, Germany. In total 92 clean mid-stream catch urine samples from patients were collected for pathogenic bacteria identification. Two mL per patients urine sample were stored at 4 °C for direct culture-independent bacteria identification and cultivation purposes. The rest of the urine samples were stored at - 80 °C after less than one hour after sample collection for subsequent metabolomics analysis. Samples were identified as infected UTI samples if the colony forming unit (CFU) number was $>10^4$ per mL of urine from patients showing typical UTI symptoms. Other standard methods for UTI diagnosis that were used are described in detail elsewhere [93]. These samples were additionally subjected to metabolomics analyses. The results for the urine of 28 healthy and 28 infected individuals served as training set for prognostic model development. After the complete analysis of these training set data, another 36 urine samples, 16 from UTI patients and 20 from healthy persons, were collected and obtained metabolomics data was used for validation purposes. They were unlabeled and randomized during the analyses and the obtained results were later evaluated in the context with the information from the clinical and SSCP analyses. The study reported in this manuscript was approved by the Institutional Review Boards of the Technical University and the involved clinic. All patients provided written informed consent for the collection of samples and subsequent analysis.

2.3.2. 16S rRNA sequence analysis for identification of the uropathogens using single strand conformation polymorphism (SSCP) analysis

The total DNA from the urine samples was extracted using the BIO 101 Fast DNA Spin Kit for the preparation of DNA from soil (MP Biomedical, Strasbourg, France). Lysis of the bacterial cells was achieved by using a FastPrep24 instrument (MP Biomedical, Strasbourg, France) for 45 s at a speed of 5.5 m/s at 4 °C. For the amplification of a 400

bp fragment of the 16S rRNA gene, which included its highly variable regions V4 (positions 519 - 536 of *Escherichia coli* 16S rRNA gene) and V5 (907 - 926 of *E.coli* 16S rRNA gene), the primers Com1 (5' – CAG CAG CCG CGG TAA TAC -3') and Com2-PH (5'-CCG TCA ATT CCT TTG AGT TT – 3') were used [94]. Prior to the single strand digestion primer Com2-PH was phosphorylated [95]. Polymerase chain reaction (PCR) reaction mixtures contained 20 µM desoxynucleoside triphosphates, 0.5 µM of each primer, 2.5 U/100 µL of DNA polymerase (Thermo Taq Polymerase, NEB, Ipswich, USA) and up to 100 ng DNA in 1x ThermoPol buffer. Reaction conditions for the PCR cycles were composed of an initial denaturation step at 95 °C for 3 min, 30 cycles of 94 °C for 50 s and 72 °C for 50 s with a final extension at 72 °C for 10 min. PCR products were purified using a Qiaquick PCR purification Kit (Qiagen, Hilden, Germany). The amplified DNA quality was checked by gel electrophoresis on a 1.5% agarose gel. SSCP profiles were generated and visualized as described elsewhere.[43] The SSCP gels were bordered by DNA molecular weight marker (Roche, Mannheim, Germany). Selected bands of the SSCP profiles were excised with a sterile razor blade. The single-stranded DNA of each band was eluted for 10 min at 95 °C in 50 µL buffer containing 500 mM ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA (pH 8.0) and 0.1% sodium dodecyl sulfate. Each DNA-fragment was amplified by PCR with the corresponding Com-primers and the resulting PCR products were purified as described above. The re-amplified DNA molecules recovered from bands of the community profile were used for DNA sequence determination [96]. Bioinformatic interpretation of obtained DNA sequences was carried out using the corresponding tools at NCBI [97]. Alignments used had a minimal sequence identity of ≥ 95%.

2.3.3. Sample preparation for LC-MS analysis

Urine samples were thawed and mixed by shaking. Creatinine concentrations in all samples were determined using Creatinine Urinary Colorimetric Assay Kit (Cayman Chemical Company, Ann Arbor, USA). Urine samples were then diluted with HPLC-grade water (Sigma Aldrich, USA) to a creatinine concentration of 500 mg/L. After all the samples were normalized based on the creatinine concentration, 300 µL of the acidified urine samples were incubated at 37 °C with shaking at 1000 rpm for 10 min. Subsequently, 100 µL of urine were then added to 300 µL LC-MS grade methanol (Sigma

Aldrich, USA) containing internal standards. Samples were further incubated at 37 °C under aeration with shaking at 1000 rpm for 15 min followed by centrifugation at 13500 rpm for 15 min. For the LC-MS analyses, 200 µL were taken, otherwise the samples were stored at - 80 °C.

2.3.4. Metabolomics measurements

The metabolite compositions of the obtained extracts were analyzed on Agilent 1290 UPLC- system (Agilent, Santa Clara, USA) in the positive and negative electrospray ionization mode. Chromatographic separation of the dissolved metabolites was performed using a ZIC®-pHILIC column 3.5 µM of particle size, 200 Å pore size, 100 x 2.1 mm (Merck Sequant, Germany), coupled to a high-resolution 6540 QTOF/MS detector (Agilent, Santa Clara, USA) in detection range of 50 to 1700 m/z at 2 GHz in extended dynamic range. The LC mobile phase was: A) 20 mM ammonium formate (Sigma-Aldrich, USA) in water (Thermo, USA) with 5% acetonitrile (Thermo, USA) (pH 9) and B) 5% acetonitrile with 20 mM ammonium formate in water with a gradient from 5% A to 35% in 0.5 min, further to 75% in 1 min, to 85% in 1.3 min, to 95% in 2 min and run until 2.3 min before shifting back to 5% at 3 min until 4 min. The flow rate was 300 µL/min with an injection volume of 1 µL.

Mass spectrometry was performed using a high-resolution 6540 QTOF/MS detector (Agilent, Santa Clara, USA) with a mass accuracy of < 2 ppm. Spectra were recorded in a mass range from 50 *m/z* to 1700 *m/z* at 2 GHz in extended dynamic range. The LC-MS data files were analyzed using the untargeted approach described elsewhere [98]. Briefly, raw data processing was done by the Molecular Feature Extraction (MFE) tool in the MassHunter Qualitative Analysis B.05.00 software (Agilent Technologies) which included background noise removal. Ions were grouped by charge state relation, isotopic distribution and the presence of adducts by using the accuracy of the mass measurements. Samples were then aligned using MassProfiler Professional (B.12.01, Agilent Technologies) using the following settings: abundance filter = 5000, minimum number of ions = 2, compound alignment parameter: RT Window = 5%, RT Window (min) = 0.25, Mass Window (ppm) = 15, Mass Window (mDa) = 2, frequency filter = remain entities that appear in at least 10% of all samples.

After the MFE list was created, recursion was the next step. The Find Compounds function of the Formula tool by MassHunter Qualitative Analysis was importing the most significant features back into MassHunter Qualitative Analysis as targeted features to improve the detection of the specific features of the samples. This feature finding improved the reliability of the procedure and lead to an improvement of the accuracy of the analysis. The generated data matrix was then normalized to internal standards using the R software package (R, Auckland, New Zealand).

Statistically significant candidate compounds were further analyzed in a targeted manner by performing MS/MS analyses on the UPLC-QTOF/MS using identical conditions. Candidate compounds were used for fragmentation with the collision energies of 0 V, 10 V, 20 V and 40 V respectively. Spectra were recorded at a rate of 3 spectra/s. The peak picking and raw data processing was performed with Mass Hunter software package (Agilent, USA). Annotation of compounds detected by LC−MS was performed by searching accurate masses of features against the data available in METLIN (www.metlin. scripps.edu), KEGG (www.genome.jp/kegg), and LIPIDMAPS (www.lipidmaps.org/) and Metabolomic Discoveries' database entries of authentic standards and IDEOM database entries through peak mass within 5 ppm mass accuracy.

Retention time prediction was applied to aid metabolite annotation [99]. The annotation of the three potential biomarkers was done by additional MS/MS analysis and these spectra were used as an input for *in silico* fragmentation software Molecular Structure Correlator (MSC) (Agilent, USA). The MSC program correlates accurate mass MS/MS fragment ions for a compound of interest with one or more proposed molecular structures for that compound. First, Molecular Formula Generator (MFG) computes all possible molecular formulas based on the MS1 information of the compounds. In addition to MFG tool of Agilent's MassHunter package, IDEOM was used for formula generation as well. Then, MSC software tries to map each observed fragment ion to the proposed structures giving each proposed structure a score. Then, MS2 spectra of the significant compounds was compared against the proposed structures and they were annotated in this manuscript based on the highest score given by the *in silico* fragmentation tool. Potential biomarkers possessed a proposed structure with a score of ≥ 95 and the candidate structures were obtained by querying a PubChem database [100].

2.4. Results and Discussion – Part I

2.4.1. Identification of the bacteria in urine samples

In this study, overall 92 urine samples were analyzed for the presence of bacteria in this study, 44 from patients with diagnosed UTI and 48 from healthy subjects. For the identification of the bacteria DNA prepared from the urine samples was subjected to SSCP analysis in combination to 16S rRNA gene sequencing. The results are summarized in Table 1.

Table 1: Overview of all identified bacterial genera found in the investigated urine samples when analyzed with PCR-SSCP method and *in vitro* cultivation. In 21 urine samples more than one bacterial species was found.

Uropathogen	**Healthy samples** n=40	**Infected samples** n=52
Escherichia	0	14
Staphylococcus	1	9
Aerococcus	0	3
Proteus	0	4
Lactobacillus	2	8
Thiobacillus	0	1
Enterococcus	2	13
Citrobacter	0	2
Prevotella	1	8
Pseudomonas	0	2
Finegoldia	0	1
Enterobacter	0	1
Veillonellaceae	0	1
Porphyromonas	0	1
Morganella	0	1
Fusobacterium	0	1

In 14 samples *Escherichia coli*, in 13 *Enterococcus*, in 9 *Staphylococcus* and in 8 *Lactobacillus* and *Prevotella* were discovered. Other genera including *Aerococcus*, *Proteus*, *Pseudomonas*, *Morganella* and *Citrobacter* were found in 1 to 4 urine samples. In a few samples of healthy persons typical commensals like *Staphylococcus* or

Lactobacillus were detected. Overall, analytical urine samples from UTI patients contained typical known urinary tract infectious bacteria. The SSCP analysis demonstrated in 21 out of 44 urine samples from UTI cases the presence of more than one pathogen bacterium.

2.4.2. Urinary metabolomics

Next, the metabolites from the urine samples were extracted and subjected to an untargeted UPLC/MS-based metabolome analysis as outlined in the experimental section. The data matrix was constructed with samples as observations and peak intensity as response variables for further analysis, accounting for 2027 peaks in total. The training data set consisted of the metabolites from the urine samples of 28 healthy and 28 infected persons and was used to develop a prognostic model. Our task was to uncover the effect of an infected status in general and the influence of certain bacteria in detail on the metabolite composition. Levels of 1783 putatively annotated metabolites in positive and 993 metabolites in negative mode were measured. Each collected sample was labelled as infected, healthy or as blank Consideration of the "blank" (extraction solvent, methanol containing internal standards) identified the confounder. Boolean infected status of a sample was used by a linear regression as an independent predictor [101].

Boolean data indicate two values, in this case the two values were "infected" and "non-infected". In order to determine which metabolites were indicative of the infection status, the relation between the metabolite level and the infection status was calculated by multiple linear regression analysis with the least squares method separately for each metabolite. The method of least squares fits a linear function to a dataset by minimizing the sum of the squares of the residuals (r_i). The residuals are errors, which occur in the difference between the calculated function ($f(x_i, \beta)$) and every point (x_i, y_i) given in the dataset,

$$r_i = y_i - f(x_i, \beta).$$

The least squares minimize the sum of all residuals r_i,

$$\min S = \ min \sum_{i=1}^{n} r_i^2 = \ min \sum_{i=1}^{n} (y_i - f(x_i, \beta))^2.$$

The regression computed the gradient between the amount of the metabolite of the infected and the healthy patient. The associated p-value indicates how significant the gradient differs to zero, so how significant the amounts of the various metabolites of the infected and non-infected differed. Since tested for multiple outcomes (each metabolite representing a single outcome), it was necessary to prevent the accumulation of false positives with each test. Multiple test corrections had to be performed controlling the family-wise error rate. For this purpose, the Holm-Bonferroni correction procedure was employed [102]. A new significance level $FWER$, Family wise error rate, was defined with the given significance level $\propto = 0.05$ and $FWER \leq \propto$.

The metabolites m_k and the corresponding p-values p_k for the gradient were sorted in ascending order with $k \in (1, \dots, N)$, with N the number of all metabolites. For the given significance level $\propto$, let k be the minimal index such that $p_k > \frac{\propto}{N+1-k}$. If such k exist, all metabolites m_{k^*} with $k^* \in (1, \dots, k-1)$ are defined as significant, all metabolites m_{k^*} with $k^* \in (k, \dots, N)$ are not significant. Using this approach, we established that the association between the infected status of the samples, regardless of the causing pathogen, and the putatively annotated metabolites heneicosanoyl-glycero-3-phospho-(1'-glycerol), thieno-(2, 3-c) pyridine and a metabolite with molecular formula $C_9H_{17}NO$ were statistically significant. Out of 1034 metabolites in the positive mode only 3 of them had $p < 0.05$ after the correction being the criterion for metabolites to be considered as potential biomarkers.

Metabolites measured in the negative ionization mode failed to reveal significance according to described criterion. The significant metabolites were used in a principal component analysis (PCA) with the training data set in order to decipher whether an unsupervised method of multivariate statistical analysis could discriminate different groups, shown in Figure 6.

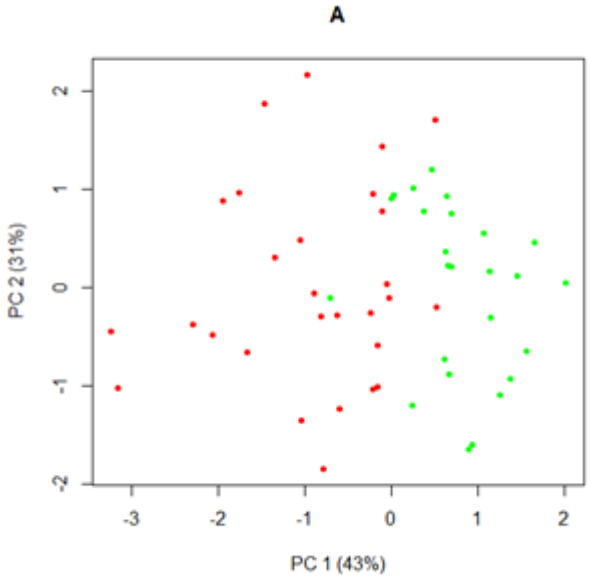


Figure 6: Scores plot of the Principal Component Analysis (PCA) prognostic regression model built with a mass spectral data derived from the from urine training set, A, and the validation set, B. colored by infectious status. Red dots indicate samples from infected patients and green dot samples from the healthy subjects. The first principal component of the training set, A, accounted for 43% while the second principal component accounted for 31% overall variability.

This PCA model yields a graphical display and offers an interpretation tool that enables investigation of the key metabolites of UTI. The scores plot of the first two components PCA model showed that the UTI group can clearly be separated from the healthy group. Finally, no obvious relation of the discovered metabolites to the identity of the infectious bacteria in the analyzed samples was observed. In agreement, all suggested potential biomarkers are most likely host-derived. Data for healthy group tended to cluster to the right part of the figure, whereas that for UTI group clustered to the left. This model provided good group separation, with first principal component accounting for 43% of overall variability while the second principal component explains 31% of overall variability.

2.4.3. Validation of putative biomarkers

For the validation of the obtained results, a second PCA model with the results from the initial training set and a novel validation set, consisting of metabolomics data for the urine of 20 healthy and 16 UTI affected persons was applied. Again, a trend separating infected samples from the non-infected samples became obvious with the first two principal components, as shown in Figure 7, indicating that combination of the new datasets with the observed data of the 3 statistically significant compounds was driving the analysis and contributed to the clustering.

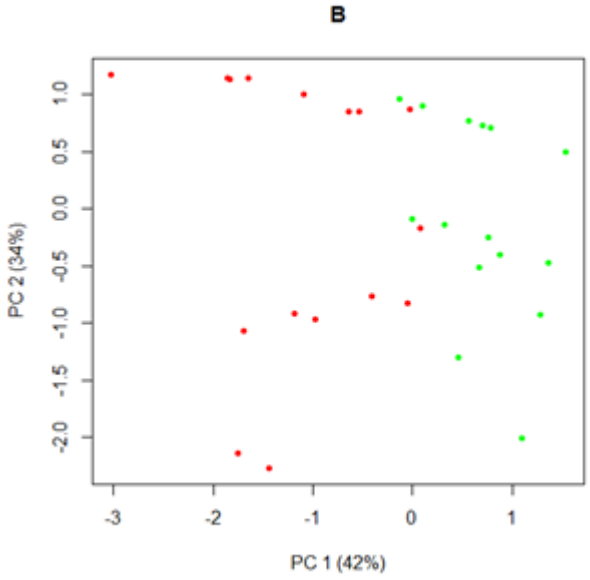


Figure 7: The first two principal components of the validation set, B, explain 76% of overall variability.

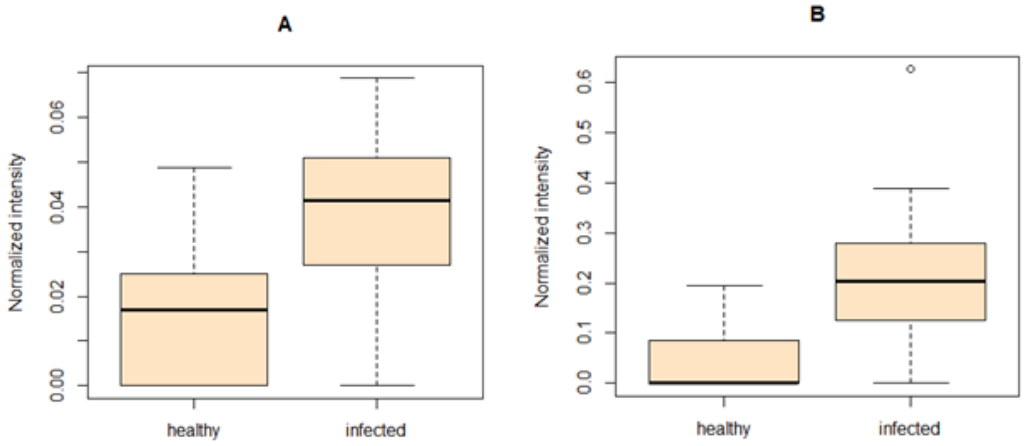
This dataset confirmed the results from the training set with two distinct visible clusters. The applied statistical model predicted 76% of the first two principal components, putatively annotated thieno-(2,3-c) pyridine, heneicosanoyl-glycero-3-phospho-(1'-glycerol), and $C_9H_{17}NO$. The concentration of putatively annotated compound, thieno-(2, 3-c) pyridine (Figure 8A) was found increased by more than 2 fold in the urine of the infected patients. This metabolite was described as a cell adhesion-inhibiting anti-inflammatory compound [103]. Thus, elevated levels in patients with UTIs are to be expected. Second putatively annotated biomarker for this syndrome is the metabolite

heneicosanoyl-glycero-3-phospho-(1'-glycerol) that was found significantly increased in the samples from the UTI patients as shown in Figure 8B.

Human urine usually contains only very small amounts of lipids. However, under certain nephrotic syndrome conditions the urinary excretion of cholesterol, cholesterol esters, triglycerides, free fatty acids and phospholipids is considerably increased [104].

Previous studies also showed that increased urinary phospholipids may indicate early aminoglycoside toxicity in patients [105].

The third statistically significant compound, $C_9H_{17}NO$ (Figure 8C), that was found increased in UTI patients and their experimental MS/MS spectra resulted with multiple possible isomers with an *in silico* fragmentation score of < 85. This indicated the *in silico* fragmentation software failed to narrow down the possible structure for the given experimental MS/MS spectra.



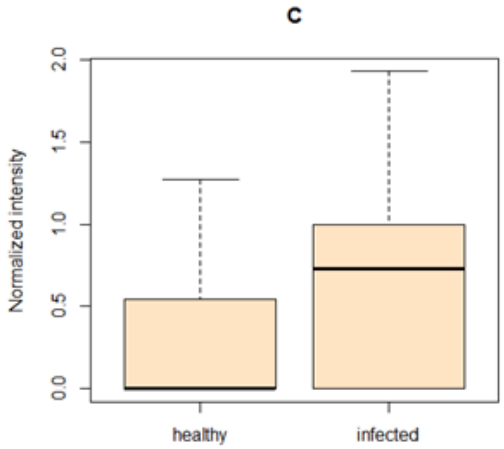


Figure 8: Box plots of the normalized peak intensities of representative putative biomarkers. (A) Thieno-(2,3-c) pyridine, (B) Heneicosanoyl-glycero-3-phospho-(1'-glycerol), (C) $C_9H_{17}NO$.

However, the metabolite with the same molecular formula was found in urine samples of patients suffering from diabetic nephropathy [106]. *In silico* fragmentation algorithms, even though they are used widely in the community and represent reliable tools for structure elucidation, may also have a false discovery rate [107]. Therefore, our next step will be the synthesis of all putatively annotated biomarkers and the comparison of their fragmentation pattern to the ones of our analyses in order to confirm the structure of these evidently important compounds.

Previously reported studies have focused on the discovery but rarely on the validation of metabolomic biomarkers. For this purpose each statistically significant metabolite of the test and validation sets, a binary logistic regression model was established to obtain the receiver operating characteristic (ROC) curve of the combinatorial biomarkers [108]. First, univariate ROC curve analyses were applied to quantify the predictive performance of each potential biomarker. The specificity and sensitivity trade-offs were calculated for each significant metabolite using the area under the ROC curve (AUC).

To further evaluate the usefulness of the putatively annotated metabolites combinations for the diagnosis of UTI, multivariate exploratory ROC analysis was performed. The AUC for thieno-(2,3-c) pyridine was 0.82 in the training set and 0.88 in the validation set, as shown in Figure 9A and B, respectively.

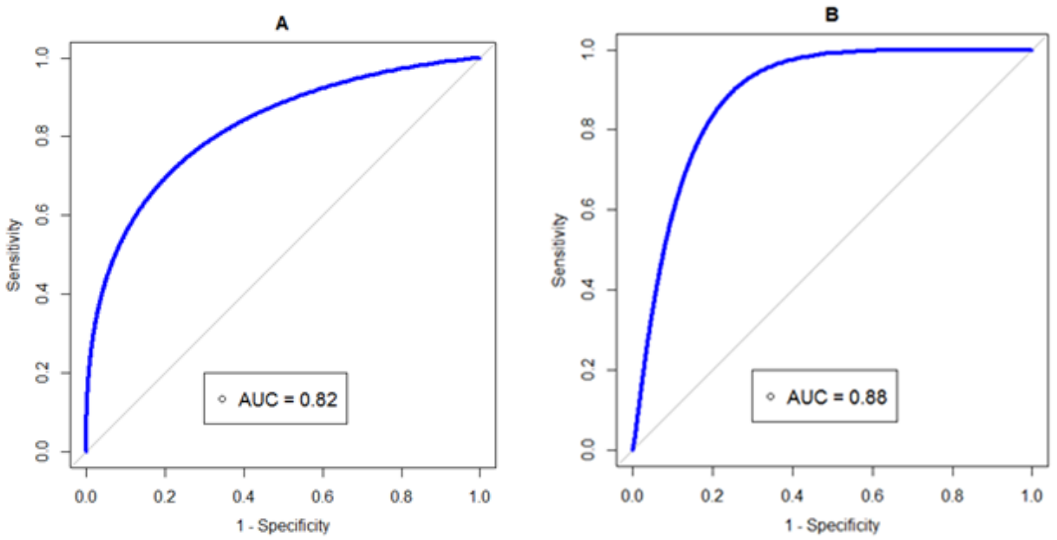
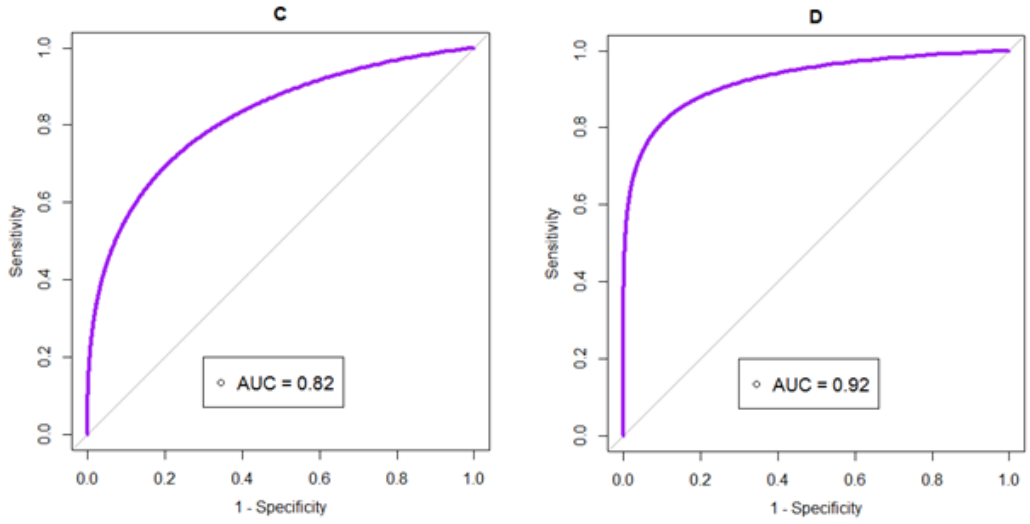
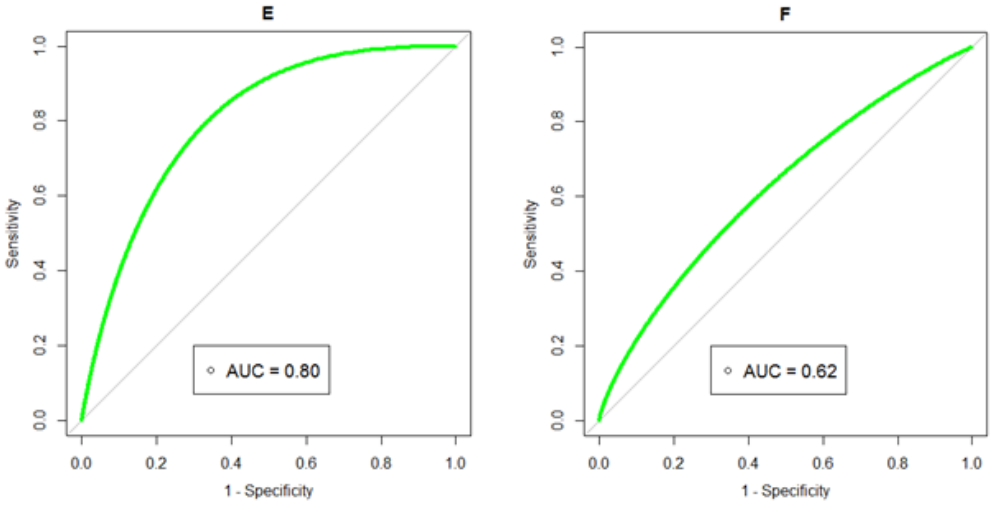


Figure 9: Receiver operating characteristic (ROC) curves of the putative biomarkers for Thieno-(2,3-c) pyridine (A) in the training set and (B) in the validation set, for Heneicosanoyl-glycero-3-phospho-(1'-glycerol) (C) in the training set and (D) in the validation set and for $C_9H_{17}NO$ (E) in training set and (F) in the validation set

The AUC for heneicosanoyl-glycero-3-phospho-(1'-glycerol) in the training set was 0.82 and 0.92 in the training set (Figure 9C and D).



The AUC for metabolite $C_9H_{17}NO$ went from 0.80 in the training to 0.62 in the validation data set, as shown in the Figure 9E and F, indicating that further validation of this metabolite is needed.



As it is known that a single biomarker is not considered adequately sensitive and specific for clinical screenings, results of multiple markers could be combined in order to accurately classify patients' diagnosis [109]. For this purpose, combinatorial ROC curve analyses of the all three putative biomarkers were performed. The AUC in the training set was 0.97 and 0.99 in the validation set (Figure 10A and B).

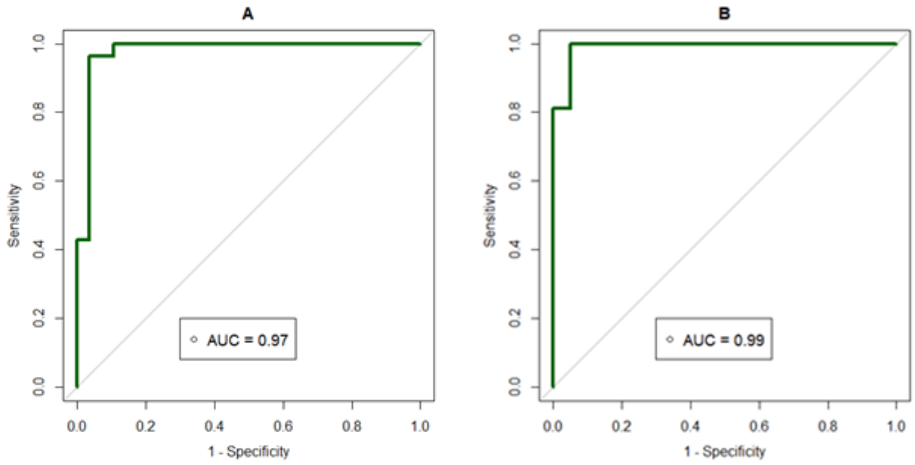


Figure 10: Receiver operating characteristic (ROC) curves of the combinatorial putative biomarkers (A) in the training set and (B) in the validation set

Metabolomics community seems to have dynamic similar approaches when discovering metabolite signatures when the reference material is missing [110]. While this information can be of great value further analysis with synthesized standards needs to follow this experimental set up in order for real life applications.

Taken together, these results indicate that the putative biomarkers, alone or in combination, can discriminate the infected samples from the controls with high accuracy and should be discussed as a powerful tool in UTI diagnostics. Characteristics of the putative biomarkers are listed in Table 2.

Table 2: Characteristics of the putative biomarkers for urinary tract infection detected in LC-MS chromatograms with the developed prognostic model

Annotation	Molecular formula	Experimental parent ion	Ionization	Mass error (ppm)	InChIKey	FC
Thieno-(2,3-c)pyridine	C7H5NS	136.0215	positive	0.675	GDQBPBMIAFIRIU-UHFFFAOYSA-N	2.3
unknown	C9H17NO	156.1378	positive	1.265	BTQGBUASSGEZMW-UHFFFAOYSA-N	2.21
Heneicosanoyl-glycero-3-phospho-(1'-glycerol)	C27H55O9P	555.3653	positive	0.279	XEIBLFIFGABXTL-IZZNHLLZSA-N	4.89

2.5. Conclusion – Part I

State of the art metabolomics was performed on urine samples from patients suffering from urinary tract infections in comparison to healthy subjects. The combination of different statistical approaches identified three host-derived compounds indicative of a urinary tract infection: 1. thieno-(2,3-c) pyridine, 2. heneicosanoyl-glycero-3-phospho-(1'-glycerol), and 3. a metabolite with the molecular formula of $C_9H_{17}NO$.

Importantly, the additional confirmation via receiver operating characteristics (ROC) curve analysis validated these findings. Found metabolites are independent of the nature of the infecting agent. Here we present a novel approach for the reliable identification of putative biomarkers in urinary tract infections which could serve for the development of inexpensive and dependable diagnostic tools.

2.6. Acknowledgement – Part I

This work was funded by Deutsche Forschungsgemeinschaft (German Research Foundation), Bundesministerium für Bildung und Forschung (BMBF, the Federal Ministery for Education and Research) and we are grateful to the Cusanuswerk (KAAD) for support.

We greatly appreciate the help of Hrvoje Torbašinović and Manuela Hische for their support with programming and modelling, Judith Lenk, Monique Thiele and Stefan Barthels for their excellent technical assistance and Metabolomic Discoveries GmbH for providing all the tools needed for this research.

Furthermore, we are very thankful for a cooperation with Urology Clinic in Braunschweig and their staff.

3 Comprehensive re-analysis of CASMI 2016 results: *in silico* MS/MS fragmentation tools must be combined with database boosting to achieve 92% accuracy – Part II

Ivana Blaženović[1,2], Tobias Kind[3], Hrvoje Torbašinović[4], Slobodan Obrenović[4], Sajjan Mehta[3], Hiroshi Tsugawa[5], Tobias Wermuth, Nicolas Schauer[2], Dieter Jahn[1], Oliver Fiehn[3,6§]

[1] Technische Universität Braunschweig · Institute of Microbiology, Braunschweig, Germany

[2] Metabolomic Discoveries GmbH, Potsdam, Germany

[3] West Coast Metabolomics Center, UC Davis, Davis, CA, U.S.A.

[4] Inovatus Ltd, Zagreb, Croatia

[5] RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa, Japan.

[6] Department of Biochemistry, Faculty of Sciences, King Abdulaziz University, Jeddah, Saudi-Arabia

[§]Corresponding author

Email addresses:

OF: ofiehn@ucdavis.edu

My contribution: design of experiments (with T.K. and O.F.); parameter optimization for all tools used; experimental procedure of MS/MS analysis; implementation of voting/consensus model; first draft of the manuscript.

3.1. Abstract – Part II

In mass spectrometry-based untargeted metabolomics, rarely more than 30% of the compounds are identified. Without the true identity of these molecules it is impossible to draw conclusions about the biological mechanisms, pathway relationships and provenance of compounds. Public libraries of tandem mass spectra are still very small in comparison to the wealth of natural products with hundreds of thousands known compounds today. The only way to address this discrepancy is to use i*n silico* fragmentation software to identify unknown compounds by comparing and ranking theoretical MS/MS fragmentations from target structures to experimental tandem mass spectra (MS/MS).

We compared the performance of four publicly available *in silico* fragmentation algorithms that participated in the 2016 CASMI challenge. We used all 312 MS/MS training spectra and the 208 challenge MS/MS spectra to improve the outcomes of each of the existing tools (MetFragCL, CFM-ID, MAGMa+ and MS-Finder). For the category 2 CASMI contest, pure *in silico* methods were compared solely based on information provided by the CASMI organizers. First, parameters were optimized by changing mass resolution to 5 ppm search windows, yielding only for MetFragCL a significant improvement from 12% to 25% correct top hits. Next, we combined the results of the four tools into consensus models by testing different combinations of tools and voting models yielding 27% correct top hits as the optimal outcome. However, metabolomics researchers would always try to rank hits by using other metadata, resources, databases and weighing factors. This approach was defined as challenge in the category 3 of the CASMI contest. We found that optimizing the use of metadata or weighing factors eventually defined the ultimate outcomes of each method. The voting/consensus method improved the original CASMI outcomes to up to 87% correct hits. In the last boost, we performed a comprehensive analysis how outcomes of the different tools could be combined and reached a final success rate of 92% for the training data and 87% for the challenge data, using a combination of MAGMa+ and CFM-ID along with MS/MS matching. Boosting without using any *in silico* tool yielded 62% correct hits showing that the use of in silico methods is still important.

3.2. Introduction – Part II

Many fields of research, from environmental analysis to forensics and biology, move towards hypothesis-generating screening approaches using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) [111, 112]. Such approach yields hundreds to thousands of signals per study, most of them of unidentified structures even after comprehensive searches of existing mass spectral libraries such as NIST, MassBank, Metlin or MassBank of North America (MoNA). Overall, tandem mass spectral databases cover less than one percent of the compound space that is covered in ChemSpider or PubChem with 50 to 90 million compounds.

As an alternative strategy for compound annotation *in silico* fragmentation software tools have been developed and are used to identify MS/MS spectra when the reference MS/MS spectrum is not available.

Such software tools include MetFragCL [92], MIDAS [76], MAGMa [113, 114], MAGMa+ [115], MOLGEN-MS/MS [116], CSI:FingerID [117], CFM-ID [118], FingerID [119], Input output kernel regression (IOKR) [120] and the MS-Finder software [121]. A number of commercial software solutions such as MassFrontier (HighChem), MS-Fragmenter (ACDLabs) or Molecular Structure Correlator (Agilent) are also available, but lack open access code or algorithm transparency.

The data for our investigation was obtained from the CASMI website (http://www.casmi-contest.org/2016/). The Critical Assessment of Small Molecule Identification (CASMI) contest was founded in 2012 to help scientists with their compound identification methods by providing community challenges and competitions [122]. For practical reasons, including the source code and model availability, error handling, batch processing capabilities and the ability to perform local database queries we only covered *in silico* fragmentation software that were used for results submitted by the CASMI 2016 deadline. We surveyed four different tools that all use different algorithms for *in silico* fragmentation, MetFragCL, CFM-ID, MAGMa+ and MS-Finder. MetFragCL retrieves candidate structures and fragments them using a bond dissociation approach and those fragments are then compared to the product ions in a measured spectrum to determine which candidate best explains the measured compound by assigning it a score that is a function of the mass to charge ratio (m/z), intensity and bond dissociation energy (BDE)

of the matched peaks, while 5 of neutral loss rules account for rearrangements [92]. CFM-ID (competitive fragment modelling) employs a method for learning a generative model of collision induced dissociation fragmentation [118]. CFM-ID can be used to assign fragments to spectra to rank the candidates, but also to predict MS/MS spectra from structures alone. MAGMa+ is a parameter optimized version of the original MAGMa software [113]. MAGMa analyses substructures and utilizes different bond dissociations. It furthermore calculates a penalty score for all the bonds that are disconnected and form a specific substructure [123]. The improved MAGMa+ version utilized a parameter optimization approach to find optimal processing parameters [115]. The MS-Finder algorithm simulates the alpha-cleavage of linear chains up to three chemical bonds and considers also bond-dissociation energies. Multiple bonds (double-, triple-, or cycles) are modelled as penalized single bonds in which hydrogens are lost (hydrogen rearrangement rules). The total score also includes mass accuracy, isotopic ratio, product ion assignment, neutral loss assignment and existence of the compound in an internal structure database [121]. First-principle quantum chemical models for spectrum prediction [124] have only been developed for electron ionization but not for electrospray collision induces dissociation tandem mass spectrometry (ESI-CID-MS/MS). The CASMI 2016 contest consisted of three categories. Category 1: "Best Structure identification on Natural Products", with 19 natural product dereplication challenges. The data for Categories 2 and 3 consist of training sets and challenge sets of 312 and 208 compounds respectively. For Category 2: "Best Automatic Structural Identification - *In Silico* Fragmentation" no other information than the *in silico* fragmentation was allowed. Category 3: "Best Automatic Structural Identification - Full Information" allowed for any type of additional information to be used, including mixed models, structure rankings and MS/MS search. In order to obtain the ground truth of performance of *in silico* fragmentation software it is important to exclude all pre-knowledge or any bias such as molecular formula lookup, database ranking, or any other means that would influence the score. Furthermore, it is important to include a large number of unknown compounds in order to improve the statistical power of the investigation. We therefore chose the 312 training and the 208 challenge MS/MS spectra for investigating the capabilities of current software to perform unbiased batch-processing of hundreds of test and validation cases. Additionally, we compare the tools' performances when more information is allowed to be used and how consensus modelling can improve results.

3.3. Materials and Methods – Part II

3.3.1. Tandem mass spectral input data

The CASMI 2016 website (http://www.casmi-contest.org/2016/) provided 312 training and 208 validation files contained MS/MS information as *.MGF file. The MS/MS spectra were acquired on a Q Exactive Plus hybrid quadrupole Orbitrap mass spectrometer (Thermo Fisher), with <5 ppm mass accuracy and MS/MS resolution of 35,000 using ESI ionization. Spectra were collected in stepped 20, 35 and 50 eV in mode. Only $[M+H]^+$ (positive) and $[M-H]^-$ ion species were available. Spectral meta-data included the ChemSpider ID, compound name, the monoisotopic mass, molecular formula, SMILES, InChI and InChIKey. Some of the candidate structures were erroneous and did not match the provided formula, SMILES or InChIKey. After contest deadline, the CASMI organizers provided all correct results for the 312 training and 208 challenge cases that were used in our evaluation.

3.3.2. Query compounds from ChemSpider

The CASMI team provided for each of the training and validation cases possible candidate lists. These compounds were obtained from ChemSpider with a ±5ppm search window and the structure files contained the ChemSpider ID, compound name, monoisotopic mass, molecular formula, SMILES, InChI and InChIKey. Because compound masses are unevenly distributed, some mass spectra yielded up to 8000 possible structure candidates within the 5 ppm mass window, whereas one mass spectrum was only associated with a single possible candidate structure, pentabromophenol. A total of 432,511 candidates were available for the training set and 258,808 candidates were obtained for the validation set (challenge set).

Each of the four software tools used different structure handling libraries or routines, hence structure conversion issues occurred. Such errors can be attributed to salt forms, isotopic elements, radical compounds and conversion issues. Each of the four tested software tools required different input formats and output formats. For that purpose, an

application was written in Java to pre- and post-processes all files. Source and result files can be found under (https://sourceforge.net/projects/in-silico-fragmentation/).

3.3.3. Software settings

3.3.3.1. MS-Finder

The MS-Finder software (version 1.70) was downloaded from the Riken institute website (http://prime.psc.riken.jp/Metabolomics_Software/MS-FINDER/index.html) and was used on a standard personal computer with a 2.50 GHz Intel Core-i7 CPU and 16 GBytes of RAM under the Windows 10 operating system. MS-Finder requires specially formatted MS^1 and MS^2 files as input. The settings are listed in Supplemental Table 1. The MS-Finder program has a resource folder where two databases are located that the software uses to rank the candidate structures. The file ExistStructureDB_vs8.esd is an internal structure lookup database and the file ExistFormulaDB_vs8.efd (comprising formula from 13 metabolomics databases) is used to prioritize generate molecular formulas. These databases were emptied in order to evaluate the pure *in silico* fragmentation performances for challenge 2 and a new database was created, analogous to the one of MS-Finder, from provided CASMI candidate files.
Both databases were opened in Notepad++ and all data except the header row was deleted and saved in the same format. Settings were adjusted to ±5 ppm mass accuracy and all compounds were processed in batch mode. Detailed information about the process can be obtained from the supplement section Supplemental document 1.

3.3.3.2. CFM-ID

The CFM-ID software (version 2.2, revision 26) was downloaded from https://sourceforge.net/projects/cfm-id/ and was used on a server with 48-core AMD Opteron 6344 processor (2.6 GHz) running CentOS Linux 7. Out of several available command line utilities, the cfm-id executable was used for this project. Given an input spectrum and a list of candidate SMILES (or InChI) as provided by CASMI, cfm-id computes a predicted spectrum for each candidate and compares it to the input spectrum. It returns a ranking of the candidates according to how closely they match. The original

CFM positive and negative models were used for the spectrum prediction, which were originally trained on data from the Metlin database. Mass tolerances of ±5 ppm were used and the Jaccard score and dot product score were applied for spectral comparisons. The better rankings produced by this comparative method were used for final evaluation. The input spectrum was repeated for the low, medium and high energies, which originally emulate 10, 20 and 40 eV CID MS/MS spectra. Additional information is contained in Supplemental document 1.

3.3.3.3. MetFragCL

The command line version of MetFragCL software (version 2.2-CL) was downloaded from https://github.com/c-ruttkies/MetFrag and was used on MacBook Pro with 2.7 GHz Intel Core i5 and 16 GB DDR3. MetFragCL needs a parameter file of specific layout as input and it contains all necessary information for the processing of a given MS/MS peak list. Parameters for fragmentation are shown in Supplemental Table 2. Candidate files were prepared with the same application used for the analysis of the results, as mentioned previously. Finally, the *in silico* fragments are matched against the query peaklist provided by CASMI. The measured peaks correspond to the charged fragments, so the matching function adds (positive mode) or removes (negative mode) a proton (1.007 Da) to the fragment mass. Additional settings are described in Supplemental document 1.

3.3.3.4. MAGMa+

The MAGMa+ software was downloaded from https://github.com/savantas/MAGMA-plus and was used on a cluster node with a 48-core AMD Opteron 6344 processor running CentOS Linux 7. MAGMa+ is an optimized version of the software MAGMa and is written as a Python wrapper script with identical command line arguments as the original MAGMa program with few changed parameters. Each candidate molecule was used to annotate the corresponding spectral tree with *in silico* generated substructures according to the algorithm published previously [123]. A Python script (process_hmdb.py) is provided that generates an SQLite .db database file from the public HMDB .sdf structures file, which is then used when running MAGMa. This script was modified to produce an analogous database file from the provided InChIs for each set of CASMI candidates. An

additional Python script was written to generate spectral-tree files required by MAGMa from the CASMI peak lists and metadata. Additional information can be found in Supplemental document 1.

3.4. Results – Part II

3.4.1. CASMI Category 2 (Best Automatic Structural Identification - *In Silico* Fragmentation Only): parameter optimization and development of voting/consensus model

Four tools that were tested in this analysis also participated in the CASMI challenge. First, parameter optimization of all tools already resulted in improved results in comparison to ones submitted to CASMI. Such parameter optimization includes using a 5 ppm window for spectral comparison. Detailed parameter setting for each tool are listed in Supplemental document. Secondly, each tool provided a ranked list of all MS/MS spectra (training and challenge) which was then used as an input for voting/consensus model resulting in new improved rankings, as described below.

3.4.1.1. *In silico* performance using the training set

Following the guidelines of the Category 2 challenge by the 2016 CASMI organizers, we evaluated each *in silico* software individually by using the best recommended settings and without secondary database rankings or use of other metadata. We utilized the 312 MS/MS spectra from the CASMI training set for parameter optimization of each tool and development of voting/consensus model. The number of compounds to be queried for each individual case ranged from less than 20 to over 8,000 compounds. The individual software tools were able identify between 10-17% of the training set as top hits (see Table 3). CFM-ID ranked the correct metabolite first in with 15% of the cases and 40% as top 5 candidates. MS-Finder ranked the correct metabolite first in 10% of cases and 27% in the top 5. MAGMa+ ranked 16% of the compounds correctly. MetFragCL was the best performing tool in our comparison placing 17% cases correctly in the top rank and 43% in the top 5 hits (Figure 11**)**.

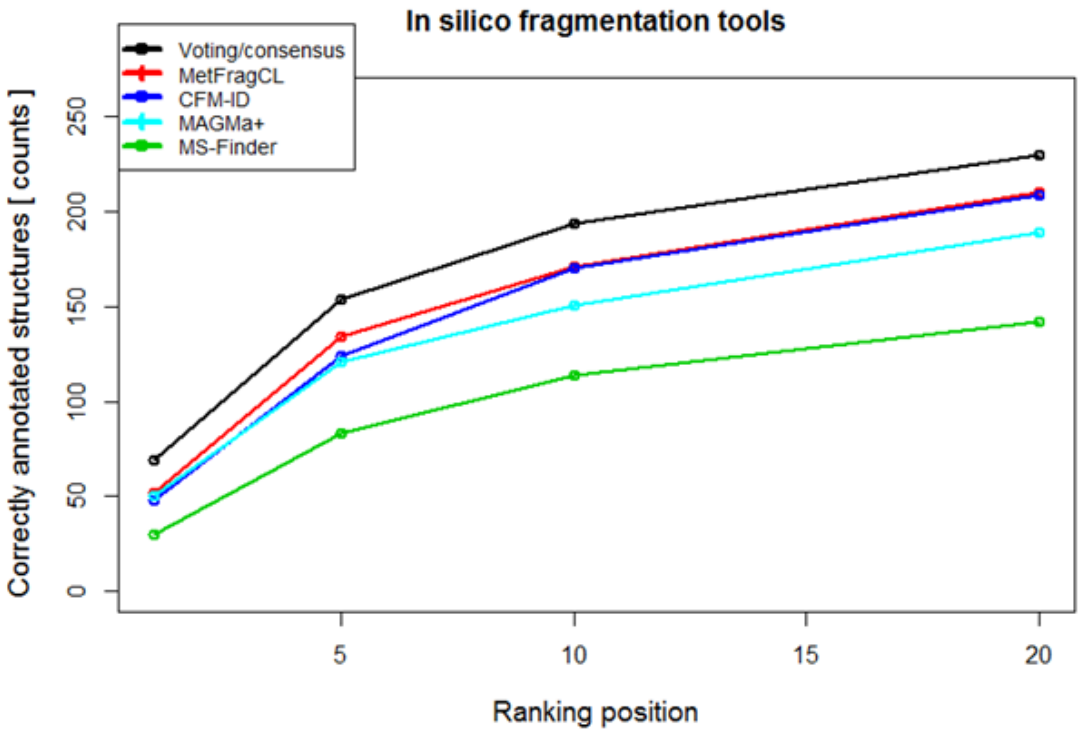


Figure 11: Comparison of all *in silico* fragmentation tools and a voting consensus model on a training data set consisted of 312 MS/MS spectra obtained in both positive and negative ionization mode

3.4.1.2. Voting/consensus model

Each software provided candidate ranking for each MS/MS spectrum from the training and challenge data set. For ranking of the structures we considered only the first block of an InChiKeys to discard enantiomeric or diastereomeric isomers. The voting/consensus model combines the ranking results of all tools and creates a new ranking system based on two criteria. The primary score of the voting/consensus model is calculated as the sum of the number of tools that successfully ranked a candidate compound. In case all four tools found the candidate structure, this primary score was four. If none of tools ranked a candidate, the score was zero. The secondary score was calculated for each candidate structure by:

$$Secondary\ score\ = Ranking\ \ (software\ A) * \omega\ (top\ 10\ software\ A)$$

where ω represents the calculated the sensitivity. Sensitivity for each software was calculated using a training data set as follows:

$$\omega = \frac{correctly\ assigned\ structures}{correctly\ assigned\ structures + falsely\ assigned\ structures}$$

Correctly assigned structures were tested with different thresholds: top rank (correct structure had to be ranked #1 by a software), top 5 (correct structure had to be within top 5 structures), top 10 and top 20. Best results were obtained with the sensitivity calculated for top 10 correctly assigned structures as shown in Table 3, and these were used for the validation set later on. By sorting the results in two levels with primary scores in descending and secondary scores in ascending order, new rankings are obtained for each candidate structure. The best voting consensus model was chosen for each experiment. The voting/consensus model was written in R script and Java. The code is freely available (https://sourceforge.net/projects/in-silico-fragmentation/).

Application of voting/consensus model to both categories is shown in the Figure 12.

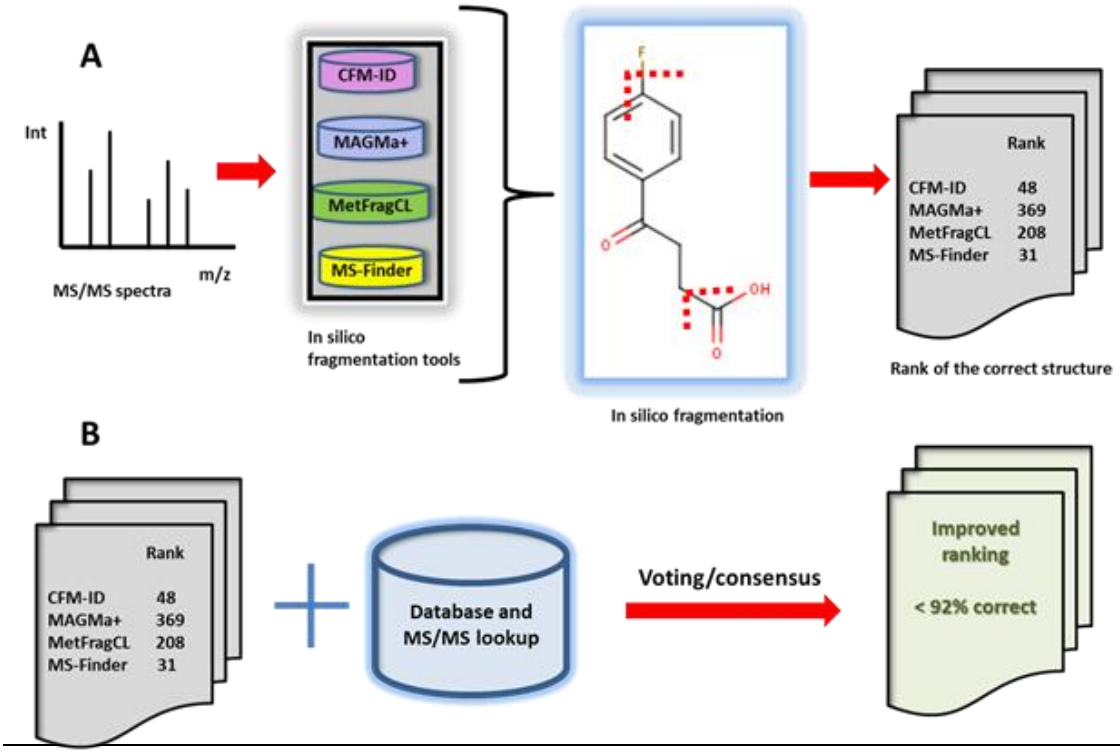


Figure 12: In silico fragmentation can be used to identify unknown MS/MS spectra by matching theoretical fragments to experimental MS/MS spectra and score all fragmented compounds.

3.4.1.3. Voting/consensus model applied to *in silico* results

Subsequently we improved overall rankings by applying the voting/consensus model as detailed in the method section. In comparison to each individual tools' results the voting/consensus model built by combination of MetFragCL and CFM-ID improved the overall results by 5%, ranking 22% cases in the top rank, 49% in the top 5 and 63% of the compounds in the top 10, as shown in Figure 11, respectively.

3.4.2. CASMI Category 3 (Best Automatic Structural Identification - Full Information): application of voting/consensus model

This category uses the same data files and candidate lists as for Category 2, but in Category 3 any form of additional information could be used (retention time information, mass spectral libraries, patents, reference count etc.). This category allows to demonstrate whether or how much additional information and meta – data can improve the results of the unknown annotation. For this category we have used database and MS/MS boosting systems for each of the tools as well as the voting/consensus model in order to further improve the rankings.

3.4.2.1. Compound database search and importance ranking

In all published research, *in silico* fragmentation tools are rarely used as stand-alone tools without searching structure databases. Nevertheless, querying public databases is important to rank *in silico* results according the occurrence or importance of compounds. For example, if a candidate result structure is contained in multiple databases, it is most likely an often observed or important molecule. Similar methods could be used for literature citation ranking or presence in target databases according to the purpose of a study (e.g. metabolomics). Two approaches were used here. First, we have combined the results of each *in silico* fragmentation tool with its membership in the local database of MS-Finder. Then we created a new ranked list for each tool and for the voting consensus model. Second, we created another voting/consensus model based on the occurrence

count in this database. The MS-Finder query structure database contains 220,213 entries sorted according to InChIKey, PubChem, exact mass, formula and SMILES. This local compound database covers structures from the most important databases including the BMDB, CheBI, DrugBank, ECMDB, FoodDB, HMDB, KNApSack, PlantCyc, SMPDB, T3DB, UNPD, YMDB and STOFF-IDENT [125].

3.4.2.2. MS/MS search settings

In addition to structure database search we have also searched MS/MS libraries as described elsewhere [126]. The NIST MS PepSearch program is a batch command-line version that is related to the NIST MS Search GUI program. Originally aimed at peptide scoring it can also be used for small molecule scoring.

Using .msp as input files, the NIST [127] and MassBank [119] MS/MS libraries were searched with a 5 ppm precursor window. Detailed parameters are listed in Supplemental Table 3. Out of 312 MS/MS spectra in the training set 276 (88,4%) had a hit in the MS/MS database with dot product score ranging from 183 (for Training-109) up to dot product score 999 (Training-029). In the challenge data set out of 208 MS/MS spectra 125 (60%) had a library coverage with dot product score ranging from 441 (Challenge-182) up to dot product score of 999 (Challenge-049). Using a training data set different cut offs of dot product scores were tested in order to determine which one results with most correctly identified MS/MS spectra. A dot product score of 400 is experimentally determined to be the best one to use on a training data set and the same cut off was then applied on a validation data set.

3.4.2.3. Sorting according to importance using ChemSpider ID

ChemSpider is a database of over 50 million chemicals from many different sources and is now maintained by the Royal Society of Chemistry [128]. The CASMI contest provided ChemSpider structures as SMILES and InChI codes together with their ChemSpider ID. ChemSpider entries are numbered by increasing numbers according to date of entry.

In general, more information is known for lower-entry number ChemSpider hits, and such compounds can be deemed more important.

High entry number ChemSpider IDs often refer to new synthetic compounds, but infrequently also to completely novel natural products. Nevertheless, low ChemSpider IDs can be given a higher relevance score during structure candidate rankings.

Ordering the candidate structures according to their ChemSpider ID alone provided correct results in many cases.

Randomization and ordering of the candidate structures allowed for another comparison of results when *in silico* fragmentation tools are not used.

3.4.2.4. Final scoring model for the validation set

For CASMI challenge 3, any combination of tools was allowed to yield rankings. Accordingly, we optimized combining *in silico* tool outcomes with presence of structure hits in compound databases and MS/MS matching into a modified voting/consensus model (see Figure 11). Finally, the optimized scoring system for the Challenge 3 was created as follows:

$$Final\ score = primary\ criterion + DB\ membership + 2 * STOFF\ ID\ membership$$

$$+$$

$$4 * MS/MS\ NISTmembership$$

Only in case of ties was the *in silico* ranking used as secondary score. Each candidate structure found in the compound databases (DB membership) was boosted once, regardless of its occurrence in the database

The STOFF-IDENT compound database [125] was used in addition and the compound occurrence in this database was boosted twice. This database was chosen for this case specifically because most of the CASMI compounds were environmentally relevant. When for example investigating metabolites the database has to be a metabolite library such as KEGG or HMDB. For each MS/MS spectrum that was found in NIST and MassBank libraries we have boosted four times to the results list of every tool. Then final rankings were obtained by sorting the sum of the scores in descending order. The higher

the final score the higher the new ranking. Overall approach boosted the overall result of each individual tool. A total of 49 different combinations were tested and the related data can be found in Table 3**.** The best voting/consensus model, built on CFM-ID and MetFragCL and it placed ~93 % correctly in the top rank and ~98 % in the top 10.

Table 3: Results of the *in silico* fragmentation performance for four publicly available software tools: MetFragCL, CFM-ID, MAGMa+ and MS-Finder. DB designates priority ranking by structure importance and MS/MS designates MS/MS library search. The 312 MS/MS spectra were the CASMI 2016 training data and the performance of each tool was compared to the voting/consensus model.

#	Tools	Top hits	Top 5	Top 10	Top 20
1	MetFragCL+CFM-ID+DB+MS/MS Voting/consensus	290	304	305	306
2	MetFragCL+DB+MS/MS	288	305	305	307
3	MetFragCL+CFM-ID+MAGMa(+)+DB+MS/MS Voting/consensus	288	304	305	307
4	MetFragCL+MAGMa(+)+DB+MS/MS Voting/consensus	287	304	305	307
5	CFM-ID+DB+MS/MS	287	304	304	306
6	CFM-ID+MAGMa(+)+DB+MS/MS Voting/consensus	287	303	306	307
7	ID-sorted+DB+MS/MS	286	306	306	308
8	MetFragCL+MS-Finder+CFM-ID+DB+MS/MS Voting/consensus	286	303	305	307
9	MAGMa(+)+DB+MS/MS	286	301	302	303
10	MetFragCL+MS-Finder+MAGMa(+)+DB+MS/MS Voting/consensus	285	302	306	307
11	MS-Finder+DB+MS/MS	285	300	302	303
12	MetFragCL+MS-Finder+CFM-ID+MAGMa(+)+DB+MS/MS Voting/consensus	284	302	305	307
13	MS-Finder+CFM-ID+MAGMa(+)+DB+MS/MS Voting/consensus	284	302	305	306
14	MetFragCL+MS-Finder+DB+MS/MS Voting/consensus	283	303	305	307
15	MS-Finder+MAGMa(+)+DB+MS/MS Voting/consensus	283	302	305	306
16	MS-Finder+CFM-ID+DB+MS/MS Voting/consensus	283	302	304	305
17	MetFragCL+CFM-ID+DB Voting/consensus	243	291	296	304
18	MetFragCL+MS-Finder+CFM-ID+DB Voting/consensus	241	290	297	301
19	MS-Finder+DB	239	284	294	296
20	MetFragCL+MAGMa(+)+DB Voting/consensus	238	290	298	301

21	MetFragCL+DB	238	290	296	301
22	MetFragCL+CFM-ID+MAGMa(+)+DB Voting/consensus	238	289	297	305
23	CFM-ID+MAGMa(+)+DB Voting/consensus	238	288	298	303
24	MS-Finder+CFM-ID+MAGMa(+)+DB Voting/consensus	237	287	298	301
25	MS-Finder+CFM-ID+DB Voting/consensus	237	286	297	300
26	MS-Finder+MAGMa(+)+DB Voting/consensus	236	288	298	299
27	MetFragCL+MS-Finder+MAGMa(+)+DB Voting/consensus	236	287	298	301
28	MAGMa(+)+DB	236	287	294	299
29	CFM-ID+DB	236	286	295	302
30	MetFragCL+MS-Finder+DB Voting/consensus	235	290	296	300
31	Randomize+DB+MS/MS	195	273	289	305
32	Randomize+DB	193	268	283	298
33	ID-sorted	143	249	267	270
34	MetFragCL+CFM-ID *in silico* Voting/consensus	69	154	194	230
35	MetFragCL+MS-Finder+CFM-ID+MAGMa(+) *in silico* Voting/consensus	67	149	185	232
36	MetFragCL+MS-Finder+CFM-ID *in silico* Voting/consensus	63	148	184	230
37	MetFragCL+CFM-ID+MAGMa(+) *in silico* Voting/consensus	60	151	189	228
38	CFM-ID+MAGMa(+) *in silico* Voting/consensus	59	137	185	222
39	MS-Finder+CFM-ID+MAGMa(+) *in silico*Voting/consensus	56	135	175	218
40	MetFragCL+MAGMa(+) *in silico* Voting/consensus	54	131	168	211
41	MetFragCL+MS-Finder+MAGMa(+) *in silico* Voting/consensus	54	130	165	212
42	MS-Finder+CFM-ID *in silico*Voting/consensus	54	114	150	201
43	MetFragCL *in silico* only	52	134	171	210
44	MAGMa+ *in silico* only	50	121	151	189
45	CFM-ID *in silico* only	48	124	170	209
46	MetFragCL+MS-Finder *in silico* Voting/consensus	47	121	146	190
47	MS-Finder+MAGMa(+) *in silico* Voting/consensus	45	110	137	194
48	MS-Finder *in silico* only	30	83	114	142
49	Randomize	4	13	27	46

3.4.2.5. Validation set performance for Categories 2 and 3

Finally, we chose the best performing methods and evaluated the 208 unknown MS/MS spectra from the validation set. This validation result mimics the approach an experienced investigator would take when identifying unknown compounds, by developing and tuning and cross-validating the algorithm on the training set and then applying the optimized parameters on the validation set. Again, each tool was used individually without any additional information and the voting/consensus model was applied using the weights calculated from the training set, as described previously (see Table 3). Here for pure *in silico* fragmentation alone MetFragCL performed best by identifying 25% of the compounds as correctly found. CFM-ID followed with 18% correctly identified compounds and MAGMa+ and MS-Finder identified less than 14% correct. When boosting the results with structure database lookup and MS/MS search the results improved tremendously to over 83% for MetFragCL and MS-Finder. The best results were obtained with 87% correct annotations for the CFM-ID and MAGMa+ voting/consensus model, as shown in Table 4.

Table 4: Results of *in silico* performance of MetFragCL, MS-Finder, CFM-ID and MAGMa+ using 208 MS/MS spectra from the 2016 CASMI contest (validation/challenge set). DB designates structure ranking importance and MS/MS library search.

#	Tools	Top hit	Top 5	Top 10	Top 20
1	CFM-ID + MAGMa + DB + MS/MS				
	Voting/consensus	181	188	194	200
2	CFM-ID + DB + MS/MS	180	191	196	199
3	MAGMa + DB + MS/MS	180	188	192	198
4	MS-Finder + DB + MS/MS	174	184	185	191
5	MetFragCL + DB + MS/MS	174	189	192	197
6	MetFragCL *in silico*	53	92	111	137
8	CFM-ID *in silico*	37	96	112	137
9	MAGMa+ *in silico*	28	71	97	116
7	MS-Finder *in silico*	22	56	78	92

3.4.8. Calculation times

We investigated a total of 520 compounds. However, each individual *in silico* tool had to process 691,319 compounds from the query database. This large number of database compounds made it challenging for a number of tools. Performance-wise MetFragCL was the fastest with only 12 hours' calculation time for the 312 training compounds, MAGMa+ needed 18 hours, whereas MS-Finder needed one day, using a regular personal computer as given in the method section. CFM-ID needed two days on a 48 CPU cluster to finish the calculation of the training set. Here additional time-out parameters can be set in the future to avoid very long computational times for individual compounds.

3.5. Discussion – Part II

3.5.1. Training set performance

Results uploaded to the CASMI contest website as well as our post-hoc tool comparison clearly shows that *in silico* algorithms alone are still far away from practical use, with only 17% of the answers correctly annotated in the training data set. Even when combining all *in silico* tools in a voting/consensus model, only 22% of the compounds were ranked as top-candidates. Importantly, even these numbers assume that 'unknowns' detected in LC-MS/MS of metabolomic or environmental studies are present as existing structures in PubChem or ChemSpider.

In fact, in actual untargeted profiling studies, many structures must be considered to be real unknowns, for example as chemical or biochemical derivatives of compounds listed in PubChem or ChemSpider databases. Few approaches exist to enumerate such database derivative structures, for example the 'metabolic *in silico* network expansion DB' (MINE)[126]. Completely de-novo spectra-to-structure calculations are yet impossible.

At current, best results were obtained when structure database importance and MS/MS search were used along with *in silico* voting/consensus models. Interestingly, each of the *in silico* tools experienced tremendous boosts, leading to 93% correctly assigned structures when combining CFM-ID and MetFrag results. Indeed, combined approaches have been successfully used in past CASMI challenges [114, 120, 129, 130]. However previous challenges did not include a large enough number of compounds for full testing. Our step-wise combinatorial multi-model approach shows a more detailed view of overall performances. Once customizable tools are available, we will extend our searches to other *in silico* fragmentation algorithms and tools such as CSI:FingerID [117] (which currently does not allow for localized database search) or the novel Input Output Kernel Regression models (IOKR) [131] (which in not yet publicly available).

3.5.2. Validation set performance and structure diversity

When the four *in silico* tools to the validation set of 208 MS/MS spectra individually and in the best selected combinations, performances observed in the training set were corroborated. With 25% correctly assigned structures MetFragCL was the best stand-alone *in silico* fragmentation tool. The voting/consensus model built on MetFragCL and CFM-ID improved the results in the top rank by only 2%. When boosting the pure *in silico* outputs by database presence and MS/MS scoring, the best individual tool to use was CFM-ID, correctly assigning 86% of the cases in the top rank. The best voting/consensus model was again the combination of CFM-ID and MAGMa+ which outperformed CFM-ID by only 1% in the top rank.

Based on PCA analysis of the molecular descriptor space we can conclude that no overfitting occurs. A number of structurally different compounds were associated only to the validation set (see Figure 13).

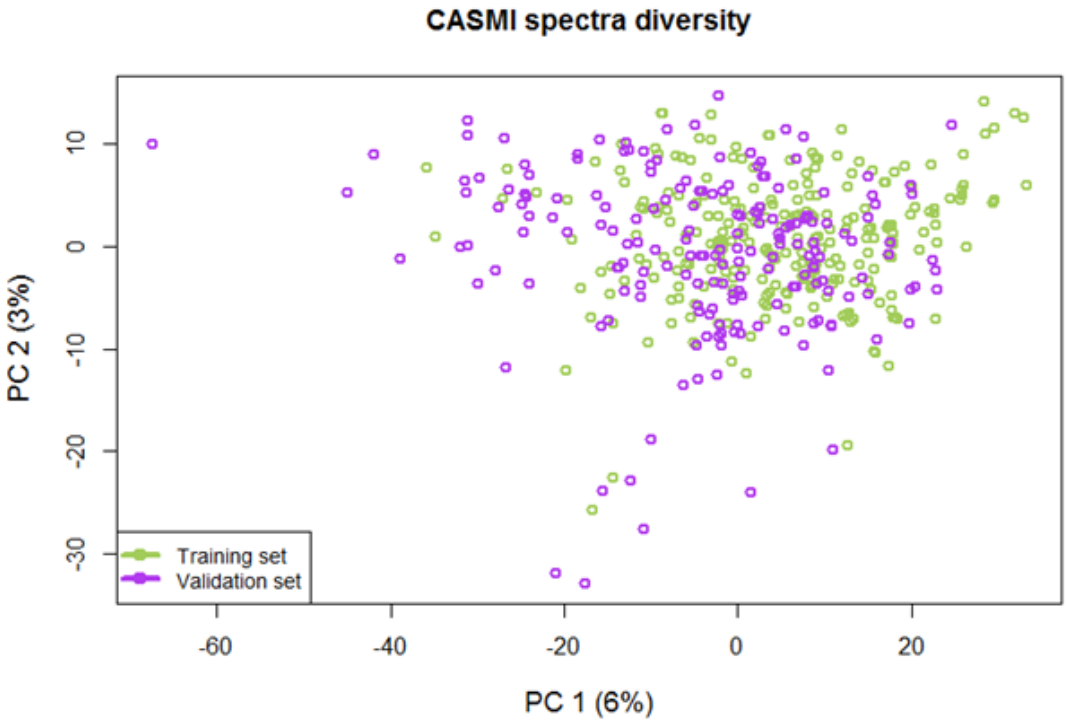


Figure 13: Principal component analysis of the molecular descriptor space from the training and validation set. Individual outliers show compounds only found in a specific data set. Overlapping dots show very similar compounds.

Lower performances of *in silico* models in the validation set compared to the training set can be explained by differences in the structural scaffolds of both sets that can be observed

when visualizing structure fingerprints by Principal Component Analysis (PCA). Indeed, structural differences between training tools (by MS/MS databases) and real world applications (in actual LC-MS/MS profiling studies) are very likely to be expected. PCA also readily separated each tool according to their ranking positions of each 208 candidates from the validation data set, accounting for 74% overall variances (see Supplemental Figure 1).

3.5.3. Voting/consensus model performance

The voting/consensus model we proposed and its secondary criterion is calculated by using the weight of each software for top 10 rankings, assuming that experts will usually rely and use the top 10 candidates proposed by a software. However, the weights could be calculated differently and it might have a better impact on the candidate ranking than the calculations we propose. The best voting/consensus model uses the results of CFM-ID and MetFragCL when only *in silico* performance is evaluated. When structural database importance and MS/MS search is added to boost the results, the best voting/consensus model is obtained from CFM-ID and MAGMa+ (see Table 3 and Table 4).

3.5.4. MS/MS data quality

The CASMI 2016 contest provided compounds were environmental xenobiotics and drugs, all covered in structure databases. About 70 MS/MS spectra had not yet been deposited in commercial or publicly available MS/MS databases. Therefore, these MS/MS spectra were not available for any software to be used as training set, rendering these spectra an excellent test for the CASMI 2016 contest to test *in silico* fragmentation algorithms. Moreover, many compounds contained fluorine atoms in their structure for which fragmentation patterns are harder to interpret. Similar to real LC-MS/MS runs, a range of challenge MS/MS spectra were sparse in the number of product ion peaks, causing *in silico* tools to fail for lack of data (Supplemental Table 4**)**. While often hundreds of isomers are retrieved per chemical formula, annotation tools must fail if too

few MS/MS product ions are generated [132]. We recommend acquiring and combining MS/MS data under multiple collision energies, or even with different mass spectrometers, for important unknowns that are detected as statistically significant in metabolomics studies.

3.5.5. Software use and performance

Offering a command-line version for *in silico* fragmentation software that is capable to run batches of test is required to process potentially thousands of unknown tandem mass spectra from profiling studies. Multi-threading and use of all CPU cores is required. However, the true challenge lies in providing tools that can be used in batch mode but are still user-friendly enough for untrained investigators. Many of these software tools operate across Windows, Linux and MacOS and require different libraries and dependencies, demanding a team environment that is skilled in cheminformatics techniques. Structure clean-up steps from the provided structure databases proved to be time demanding, involving tasks such as removing counter ions, adduct salts, or isotopes. Offering web-based research tool is recommended. However, data transfer over the web is often forbidden in industrial environments and is also prone to network errors and server outages. Because almost all mass spectrometry vendors use Windows platforms we further recommend providing Windows-based tools for *in silico* fragmentations.

For individual performance checking it is also useful to investigate each individual result with graphical user interfaces. Here MS-Finder provides a convenient desktop solution for Windows.

3.5.6. Continuous integration and design of experiment

Continuous integration refers to a concept in software technologies where multiple software tools are continuously compiled to improve results and to fix errors. This strategy includes test-driven developments were the software has to pass specific test units, in order to catch implementation errors. The Design of Experiment (DOE) is a long

standing concepts were the whole parameter space is scanned to find the best input parameters. Our voting/consensus software uses output results of all integrated *in silico* tools. It would be prudent to test formats as well as output accuracies whenever new software versions are released.

Specifically, it should be tested whether overall voting/consensus results are actually improved or getting worse by new software releases.

Our newly developed voting/consensus model software can automatically evaluate hundreds of optimization models and to report overall outcomes or top hits, top ten hits and the specificity of the model. Our software is suited to be extended to include more *in silico* software tools, with output statistics to be modified to calculate additional statistical figures of merit. Such extensions require that *in silico* software tools are publicly available in a useable form so the results can be independently validated.

One could imagine that future CASMI contests are completely run automatically by software, preventing errors and individual interventions.

3.6. Conclusion – Part II

In silico algorithms for structural fragmentation of compounds are still in early development. In many cases, existing tools only cover simple fragmentations, but not the more complex rearrangement reactions [132]. Once more and more MS/MS spectra are becoming available and corresponding structural diversity increases, these can be used to train and optimize *in silico* algorithms which will lead to better performance.

Pure *in silico* algorithms only identified 17% of the compounds correctly. The voting/consensus model slightly improved the results to 22%. Once the database and MS/MS search was added the algorithm was able to correctly identify 87-93% of the unknown compounds as the first hit, showing the importance of ranking and database queries. These results confirm that voting/consensus models can be used for real-world applications. Our software will also allow for automatic testing and performance tuning without user interaction.

The true challenge is presented by the identification of "unknown unknown" compounds that are not yet covered in compounds databases or that are computationally derived as chemical or enzymatic derivatives. Here classical experimental ways of structure elucidation including compound purification and subsequent NMR, UV and MS will play a role to elucidate the correct isomer structure.

3.7. Acknowledgement – Part II

Funding for T.K. and O.F. was supported by NSF MCB 1139644, NIH P20 HL113452 and U24 DK097154.

We greatly appreciate the help of IT team of UC Davis: Sajjan S.Mehta, Gert Wohlgemuth and Diego Pedrosa.

This work was funded by Deutsche Forschungsgemeinschaft (German Research Foundation), Bundesministerium für Bildung und Forschung (BMBF, the Federal Ministery for Education and Research) and we are grateful to the Cusanuswerk (KAAD) for support.

3.8. Supplemental material – Part II

Comprehensive re-analysis of CASMI 2016 results: *in silico* MS/MS fragmentation tools must be combined with database boosting to achieve 92% accuracy

Ivana Blaženović[1,2], Tobias Kind[3], Hrvoje Torbašinović[4], Slobodan Obrenović[4], Sajjan Mehta[3], Hiroshi Tsugawa[5], Tobias Wermuth, Nicolas Schauer[2], Dieter Jahn[1], Oliver Fiehn[3,6§]

[1]Technische Universität Braunschweig · Institute of Microbiology, Germany, Braunschweig

[2]Metabolomic Discoveries GmbH, Potsdam, Germany

[3]UC Davis Genome Center – Metabolomics, Davis, CA, U.S.A.

[4] Inovatus Ltd, Zagreb, Croatia

[5]RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa, Japan.

[6]Department of Biochemistry, Faculty of Sciences, King Abdulaziz University, Jeddah, Saudi-Arabia

[§]Corresponding author

Email addresses:

OF: ofiehn@ucdavis.edu

3.8.1. Methods

The CASMI 2016 website (http://www.casmi-contest.org/2016/) provided files contained MS/MS information as *.MGF file as well as structures for each of the challenge and validation sets. That included the ChemSpider ID, compound name, the monoisotopic mass, the molecular formula, SMILES, InChI and InChIKey.

Tools that were used for this research, information how to install and use them are presented below.

3.8.2. Software settings

3.8.2.1. MS-Finder

MS-Finder software and its documentation was taken from its website http://prime.psc.riken.jp/Metabolomics_Software/MS-FINDER/index.html.

MS-Finder program has a resources folder where two databases are located and that the software uses to rank the candidate structures (ExistStructureDB_vs8.esd) and to generate molecular formulas (ExistFormulaDB_vs8.efd). These databases have to be emptied in order to evaluate the pure *in silico* fragmentation performances. Under normal circumstances this is not required. Both databases were opened in Notepad++ and everything except the header row was deleted and saved in the same format. Unknown structures were imported as user-defined db by merging all candidates .csv files provided by CASMI using the following steps: a) Windows Start Button. b) Type cmd and hit enter ("Command Prompt" in Windows 10). c) Go to the folder with the CSV files (for help how to do that enter "help cd"). d) Type copy *.csv all.txt and hit enter to copy all data in the files into all.txt. e) Type exit and hit enter to close the DOS window. f) Open Excel. g) Click File Open to open all.txt file that was saved in the same folder where the candidate .csv files are. h) Choose Delimited. i) Next. j) Check Comma. k) Finish. l) Remove all the headers from newly created file and change the structure of the table according to the file located in the MS-Finder Local DB example folder. m) Save as .txt file. Now, user-defined database is ready to use. Once all the settings were applied,

databases and MS/MS data have been prepared, in the Analysis tab Compound annotation (batch job) was selected followed by Batch job settings set to both processes (formula finder and then structure finder) and top N hits was set to 12176 for the training and to 8555 for the challenge set (this corresponds to the candidate files with the highest number of possible structures provided by CASMI, to ensure that all given structures are considered and ranked by the software). Detailed settings are provided in the Supplemental Table 1.

Supplemental Table 1: Settings used for open source *in silico* fragmentation tool MS-Finder

#	Settings	YES/NO or other information
1	Formula finder: Lewis and Senior check	yes
2	Isotopic ratio tolerance	20%
3	Element ratio check	Common range (99.7%) covering
4	Element probability check	yes
5	Mass tolerance type	ppm
6	Mass tolerance (MS1)	5 ppm
7	Mass tolerance (MS2):	10 ppm
8	Element section (O, N, P, S, F, Cl, Br, I, Si):	yes
9	Result cut off	12176 for training and 8555 for challenge data set.
10	Structure finder: *In silico* MS/MS fragmenter setting: Tree depth	2
11	Relative abundance cut off	1%
12	Result cut off	12176 for training and 8555 for challenge data set.
13	Local Databases	User-defined db
14	MINE database	Never use it
15	PubChem online setting	Never use it

3.8.2.2. MetFragCL

Command line version of the tool and its documentation was downloaded from its website and used on a Mackintosh http://c-ruttkies.github.io/MetFrag/projects/metfragcl/. The settings that were used to run the analysis for both sets, training and validation, are listed in the Supplemental Table 2**.**

Supplemental Table 2: Setting used for open source *in silico* fragmentation tool MetFragCL

#	Settings	Information
1	PrecursorIonMode	1 (1 for positive and -1 for negative ionization mode)
2	IsPositiveIonMode	True (False for negative ionization mode)
3	FragmentPeakMatchAbsoluteMassDeviation	0.001
4	FragmentPeakMatchRelativeMassDeviation	5
5	MaximumTreeDepth	2
6	MetFragPostProcessingCandidateFilter	InChIKeyFilter
7	Adduct type of the precursor	[M+H]+ for positive and [M-H]- for negative ionization mode

3.8.2.3. MAGMa+:

A Python script (process_hmdb.py) is provided that generates an SQLite .db database file from the public HMDB .sdf structures file, which is then used when running MAGMa. This script was modified to produce an analogous database file from the provided InChIs for each set of CASMI candidates. An additional Python script was written to generate spectral-tree files required by MAGMa from the CASMI peaklists and metadata. The docker images and associated docker files that were used with detailed settings can be found here https://hub.docker.com/r/ssmehta/magma-plus/. Command line to start the run is: docker run -it -v /path/to/magma-plus/supplement:/data ssmehta/magma-plus:v1.0.0.

Once the 4 .sdf files were generated, one for positive and negative mode as well as for training and challenge data sets, they were now ready to be converted into MAGMa+ structure database file. MAGMa+ parameters that we used include -i (ionization mode: -1 for negative and 1 for positive mode), -p (maximum relative ppm error), -q (maximum absolute m/z error in Da), -c (minimum intensity of MS1 precursor ion peaks to be annotated), -d (minimum intensity of fragment peaks to be annotated, as percentage of base peak) and. The following settings were applied: read_ms_data -i -1 /1, -p 10 , -q 0.01, -c 0, -d 0 and -s hmdb.

3.8.2.4. CFM-ID

The original CFM positive and negative models were used for the spectrum prediction, which were trained on data from the Metlin database. Mass tolerances of 5 ppm were used and the Jaccard score and DotProduct score were applied for spectral comparisons, the better rankings produced by this comparative method were used for final evaluation. The input spectrum was repeated for the low, medium and high energies, which originally emulates 10, 20, 40 V CID, however this information was not available. The docker images and associated docker files that were used and the detailed settings can be found here https://hub.docker.com/r/ssmehta/cfmid/ and the command line to run the tool and reproduce all the results is: docker run -it -v /path/to/cfmid/supplement:/data ssmehta/cfmid:r25.

3.8.2.5. MS/MS database search

The NIST MS PepSearch program is a batch command-line version that is related to the NIST MS Search GUI program. The input files were .msp files and the NIST and MassBank MS/MS libraries were searched with a 5 ppm precursor window. Detailed parameters are listed in Supplemental Table 3**.**

Supplemental Table 3: Detailed parameters used for MS/MS search using the NIST MS PepSearch program

#	Settings	Information
1	Presearch mode	standard
2	m/z limits	Min. = 0 and Max. = 2000
3	Search tolerance settings	Precursor ion tolerance, m/z units 0.005
4	Ignore peaks around precursor	yes
5	Fragment peak m/z tolerance	0.5
6	Min. match factor (MF) to output (0-999)	1
7	Min. peak intensity (1-999)	1
8	Max. number of output hits	1
9	Show spectra without matches	yes
10	Include Hit-Unknown precursor m/z difference	yes
11	Include m/z in the output	yes
12	Output the input spectrum number	yes
13	Set program priority above normal	yes
14	Use number of replicates	no
15	Calculate rev-dot	yes
16	Q-TOF	no
17	Load libraries in memory	no

3.8.2.6. CSI: FingerID

CSI: FingerID combines computation and comparison of fragmentation trees with kernel methods. Kernel denotes a similarity coefficient for either MS/MS spectra or fragmentation trees and is used for the prediction of molecular properties of the unknown compound. Currently it is not possible to modify the candidate database locally thus we did not compare CSI: FingerID with the other tools in the main manuscript. However, it is of interest to see how it performs on the same data set we have incorporated the results in the supplement. CSI: FingerID was downloaded as a 64 bit GUI version from https://bio.informatik.uni-jena.de/software/sirius/ and was used on Windows. Settings are listed in the Supplemental Table 4**.** With these settings CSI:Finger ID was able to correctly annotate 140 MS/MS spectra of the validation set or 67,30%. However, since it was not possible to customize the local database in order to test the *in silico* fragmentation possibilities we did not compare these results to other tools.

Supplemental Table 4: Parameter setting for *in silico* fragmentation tool CSI:FingerID

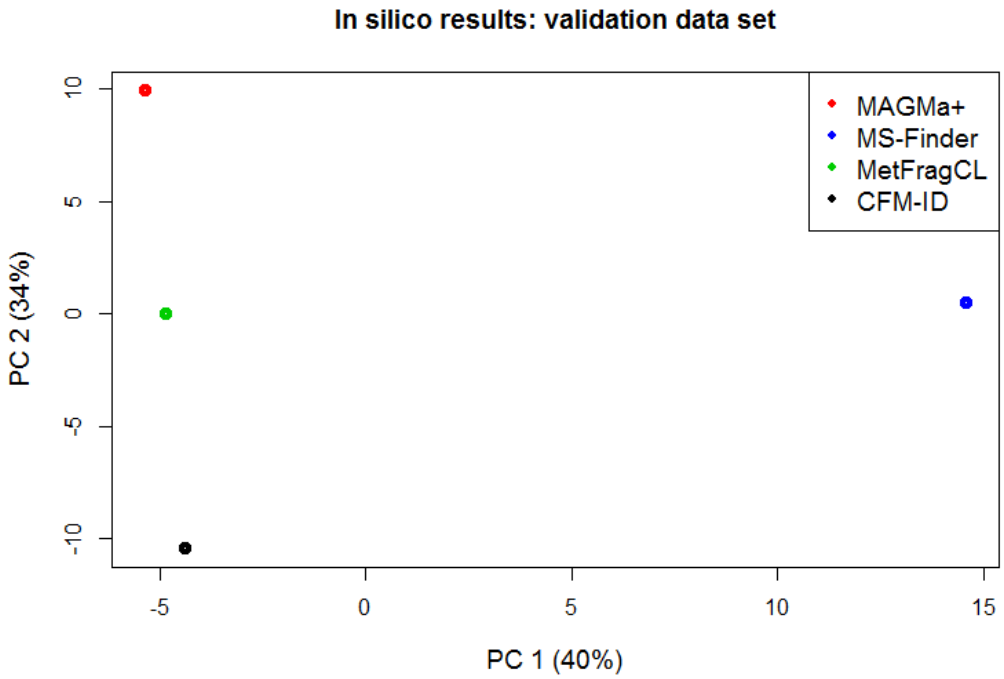
#	Settings	Information
1	File type	.mgf
2	Elements beside CHNOPS	Br, Cl, F, I
3	Search tolerance settings	Precursor ion tolerance, m/z units 0.005
4	Ionization	[M-H]- and [M+H]+
5	Instrument	Q-TOF
6	ppm	5
7	Consider	PubChem formulas
8	Search with CSI:FingerID	PubChem
9	Export	CSI:FingerID results

Supplemental Table 5 lists all the compounds that every tool we tested had problems with.

Supplemental Table 5: List of the compounds where every tool performed poorly. Ranking position calculated by each tool per correct solution is shown.

Correct solution	**Ranking position of the correct compound given by *in silico* fragmentation tool**			
InChIKey (1st block)	**MetFragCL**	**MS-Finder**	**CFM-ID**	**MAGMa+**
ZZORFUFYDOWNEF	137	541	75	308
YHQDZJICGQWFHK	98	184	205	119
BWHOZHOGCMHOBV	119	182	125	156
SYELZBGXAIXKHU	234	105	85	144
IRCMYGHHKLLGHV	151	1161	181	460
LCGTWRLJTMHIQZ	123	111	124	129
RJGDLRCDCYRQOQ	169	186	57	210
AFYCEAFSNDLKSX	123	366	142	227
VHBFFQKBGNRLFZ	213	118	53	212
CWJSHJJYOPWUGX	576	24	125	341
WWYNJERNGUHSAO	113	71	73	193
VVBLNCFGVYUYGU	1031	5460	644	1825
FRQMUZJSZHZSGN	177	116	51	271
MMBILEWCGWTAOV	361	253	45	281
CHIFOSRWCNZCFN	79	675	1160	1238

Supplemental Figure 1 shows PCA also readily separated each tool according to their ranking positions of each 208 candidates from the validation data set, accounting for 74% overall variance.



4 The metabolic response to voluntary activation of the sympathetic nervous system and attenuation of the innate immune response in humans

Ivana Blaženović [†, ‡ ,*], *Jelle Zwaag*[¥ ,*], *Daniel Stoessel* [‡,*], *Josephine M Worseck*[‡], *Nicolas Schauer*[‡], *Dieter Jahn, Peter Pickkers*[¥, §], *Matthijs Kox*[¥, §]

[†] Institute of Microbiology, Technische Universität Braunschweig, Germany

[‡] Metabolomic Discoveries GmbH, Potsdam, Germany

[¥]Intensive Care Medicine, Radboud university medical center, Nijmegen, the Netherlands

[§]Radboud center for infectious diseases, Nijmegen, the Netherlands

* These authors contributed equally to this work

My contribution: design of metabolomics experiments (with D.S. and N.S.); project coordination; data exploration and statistical analysis (with M.H. an D.S.); literature research and data interpretation; first draft of the manuscript.

4. 1. Introduction – Part III

Activation of the sympathetic nervous system has profound effects on metabolism, predominantly mediated through release of catecholamines such as epinephrine [133]. Likewise, there exists extensive interplay between inflammation and metabolism, and increasing interest in the use of metabolic fingerprints as prognostic tools in inflammatory or infectious diseases [134]. Recent advancement of metabolomic profiling, such as reversed phase chromatography coupled to electrospray mass spectrometry (ESI–MS), allow the detection of highly polar compounds that also occur in blood plasma[135]. When applied to biofluids in the field of medicine [136], the profiling of metabolites has enabled investigation of the full array of metabolites in the plasma, yielding an instantaneous and comprehensive snapshot of the "metabolic status" of a subject or patient.

We have recently showed that the sympathetic nervous system can be voluntarily activated through a training program consisting of meditation, exposure to cold, and breathing exercises, which resulted in attenuation of the innate immune response [137]. Trained healthy volunteers were able to profoundly increase their enogenous adrenaline levels, which resulted in suppression of the immune response during experimental human endotoxemia (a well-characterized, standardized and reproducible model of systemic inflammation elicited by administration of lipopolysaccharide [LPS]) [137]. However, it is unknown to what extent the studied intervention affects metabolic processes and if these play a role in the observed effects.

Herein, we have used an established metabolomics profiling platform [99] to investigate the metabolic response in subjects that were trained in the behavioral intervention and underwent experimental human endotoxemia.

4.2. Materials and Methods – Part III

4.2.1. Subjects and experimental design

Metabolomic profiling was performed on plasma samples obtained in a previously published parallel randomized controlled study registered at ClinicalTrials.gov as NCT01835457 [137]. The study protocol is described in detail elsewhere [137].

Briefly, after approval from the local ethics committee of the Radboud university medical center (CMO 2012/455), 24 healthy nonsmoking male volunteers with a median age of 22 years (range 19-27) provided written informed consent and were included in the study. All study procedures were in accordance with the Declaration of Helsinki, including current revisions, and Good Clinical Practice guidelines. Subjects were screened before the study and had a normal physical examination, electrocardiography, and routine laboratory values. The subjects were randomly allocated to the trained group (n=12) or the control group (n=12) by the opening of sealed envelopes prepared by unblinded staff not involved in the study, shown in Table 5.

Table 5: Subject demographic characteristics. Parameters were measured during screening visit. BMI, body mass index; HR, heart rate; MAP, mean arterial blood pressure. Data are presented as median (range)

Parameter	**Control group, *n*= 12**
Age, y	22 (19–27)
Height, cm	185 (179–189)
Weight, kg	78 (65–91)
BMI, kg/m2	23 (20–27)
HR, beats/min	61 (40–75)
MAP, mmHg	94 (78–105)

The trained group underwent a 10-day training program provided by Dutch individual Wim Hof, which consisted of three main elements: meditation, exposure to cold, and breathing techniques (see [137] for a detailed description). After completion of the training, subjects underwent experimental human endotoxemia at our intensive care research unit, consisting of administration of an intravenous bolus of 2 ng/kg of US Standard Reference *Escherichia coli* endooxin (*E. coli* O:113 [LPS], Clinical Center Reference Endotoxin; National Institutes of Health, Bethesda, MD, USA). The experimental human endotoxemia protocol is provided in detail elsewhere [137]. The control group did not undergo any training procedures and also underwent experimental human endotoxemia. Subjects in the trained group started practicing the learned breathing techniques half an hour before LPS administration until two-and-a-half hours afterwards. The control group did not practice any techniques throughout the endotoxemia experiment.

4.2.2. Sample preparation for LC-MS analysis

Ethylenediaminetetraacetic acid (EDTA) anti-coagulated blood was obtained one hour before LPS administration (T= minus 1 h), and at T= 0, T= 1, T= 2, T= 4 and T= 8 h. Blood was immediately centrifuged at 2000 g for 10 minutes at 4 °C after which plasma was stored at –80 °C until analysis. Sample extraction was described previously [99] with some modification. Briefly, 50 µL of plasma was mixed with 450 µL of 90% (v/v) methanol containing internal standards and incubated for 15 min at 37 °C with 1150 rpm. Precipitated proteins were separated from the extract by centrifugation for 12 min at 15 000 rpm. Supernatant was transferred into micronical vials and was stored at -80 °C until further analysis.

4.2.3. LC-MS analysis

Modified reversed-phase chromatography in combination with high resolution mass spectrometry (HRMS) was employed in this study. Samples were analyzed on an Agilent 1290 UPLC system (Agilent, Santa Clara, USA) with a Discovery HS F5-3 column (15cm

x 2.1mm, 3µm, Supelco, Sigma Aldrich) coupled to a high-resolution 6540 QTOF/MS Detector (Agilent, Santa Clara, USA) operated in both positive and negative ESI mode in a detection range of 50 to1700 *m/z* at 2 GHz in extended dynamic range. The LC solvent consisted of 95% 10 mM ammonium formate with 0.1% formic acid and 5% acetonitrile (A) and 95% acetonitrile with 5% 10 mM ammonium formate with 0.1% formic acid (B) with a multi-step gradient with 5% B from 0-0.1 min, then to 35% B at 1.5 min, to 95% B at 2.05 min kept constant until 3.2 min, to 5% B at 3.21 min and washing until 4.3 min with 5% B. The flow rate was kept constant at 700 µL/min from 0 min to 2.2 min, increased from 2.2 min to 2.5 min up to 900 µL/min until 3.2 min. The flow rate was decreased from 900 µL/min to 800 µL/min from 3.2 min to 3.1 min and constant until 3.7 min were the flow rate was changed to 700 µL/min.

The run time was 4.3 min, 1 µl of sample was injected and the column heated to 40 °C. The DualAJS ESI source was set to the following parameters: gas temperature 200 °C, drying gas 8 L/min, nebulizer 35 psig, sheath gas temp: 350 °C, sheath gas flow 11 L/min, VCAp 3500 V and nozzle voltage of 0 V. Online calibration of the instrument was performed throughout the data acquisition using Agilent ES-TOF Reference Mass Solution Kit.

4.2.4. Cytokine and catecholamine determination

Plasma concentrations of the cytokines TNFα, IL-6, IL-8, and IL-10, and catecholamines norepinephrine, epinephrine, and dopamine concentrations were determined as described previously [137].

4.2.5. Raw data processing and data analysis

Chromatograms were generated by the LC-MS instrument in .d format. Raw data was converted to mzXML and chromatogram peaks extracted XCMS [138] (v1.42.0**)** which was optimized by using the IPO R package [139], as shown in Table 6 with following settings:

Table 6: Parameter settings used for raw data processing

Parameter	Value
peakwidth	c(10, 70)
ppm	20
sntresh	10
mzdiff	0.0034
prefilter	c(3, 100)
noise	100
gapInit	0.8448
gapExtend	2.0544
bw	5
mzwid	0.015
minfrac	0.5
max	50

All further analysis was done in R programming language [140]. IDEOM software ((http:// mzmatch.sourceforge.net/ideom.php) [141] was used to eliminate noise and for putative peak annotation by exact mass within ±10 ppm against Metabolomic Discoveries in house metabolite library [98] in negative and positive ESI mode, respectively. Retention time prediction was applied to aid metabolite annotation.

For biological interpretation and visualization of the annotated metabolites iPath2.0, a web-based tool (http://pathways.embl.de), was utilized [142]. It provides an overview of important regulatory pathways directly connected to the annotated metabolites with identifiers obtained from publicly available databases such as KEGG [143] and HMDB [144] allowing a more general overview and hypothesis generation.

4.3. Results – Part III

In this study we aimed to further investigate our previous reporting by using the metabolomics platform and to identify metabolomics response in human plasma obtained from subjects that were trained in the behavioral intervention and underwent experimental endotoxemia. Snap-shot of global metabolomics profiles was obtained through untargeted analysis by UPLC-MS platform resulting in 973 detected features in both positive and negative ionization mode. As unknown metabolites are a major bottleneck in metabolomics [145] we were able to annotate 228 detected features including lipids, amino acids, alkyl-amines, carbohydrates, flavonoids and xenobiotics. Complete list of the global metabolomics profiling is provided in the appendix.

Principal component analysis (PCA) was calculated to illustrate the variances between the different sample groups. The colored ellipses describe the 50% confidence interval of the normal distribution for each group in the two-dimensional space of the PCA. For graphical representation in a heat map, the data is arranged in a matrix, in which each metabolite is depicted in one row, each sample or group is depicted in one column, and (mean) metabolite abundances in a specific sample / group are illustrated by color. Each metabolite's abundances are centered and scaled. Firstly, centering is done by subtracting the total metabolite's mean. Consequently, samples/groups with an average abundance of this metabolite get the value zero and appear white, an abundance below average get negative values and appear blue, an abundance above average get positive values and appear red. Secondly, scaling is done by dividing the centered metabolite by its standard deviation. The resulting values are in a similar range for all metabolites. Thus, the variance within a metabolite becomes comparable with other ones.

Finally, the rows and columns are rearranged in such way that rows or columns with similar profiles are grouped. To do so, similarity is determined by a hierarchical clustering, as shown in Figure 14.

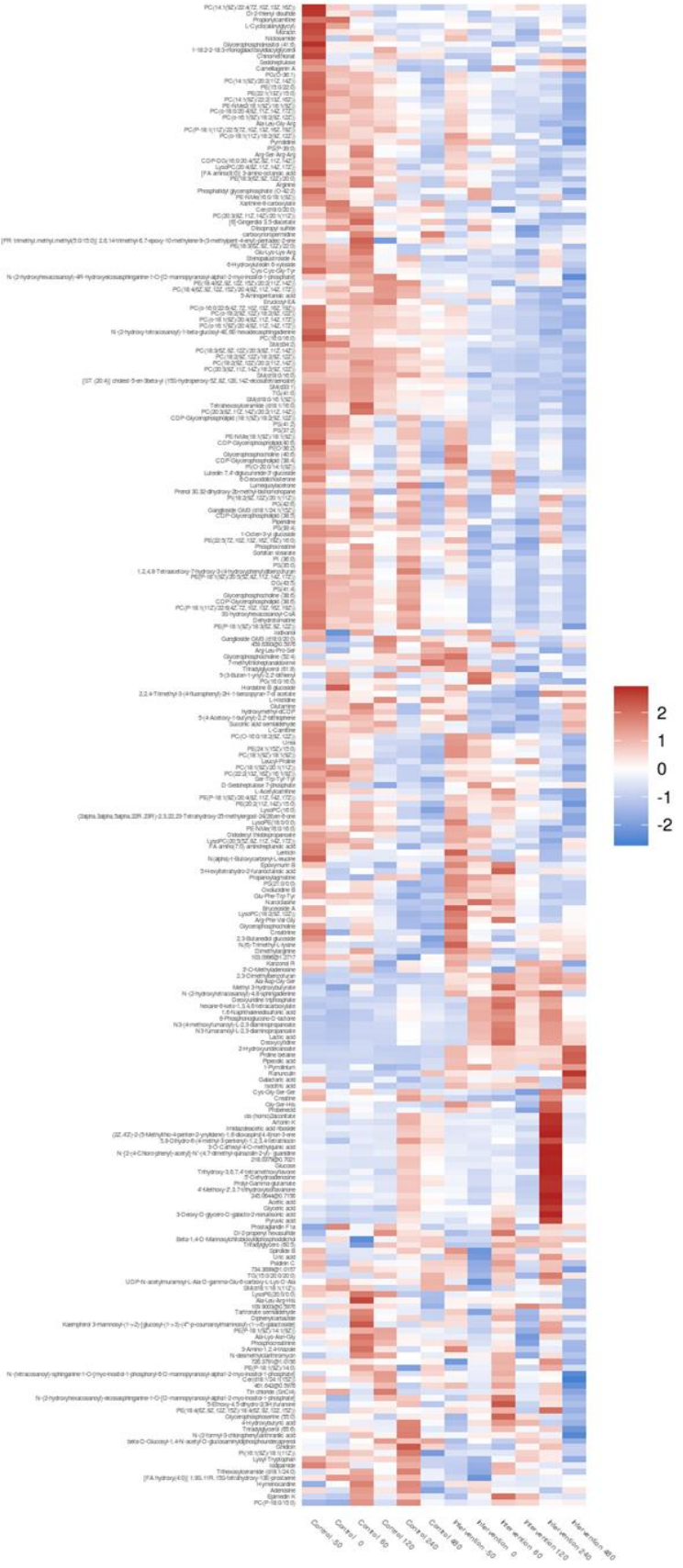


Figure 14: Heat map describing the highest differences between the trained individuals and control group from T= - 1 hour up to T= 8 hours, during experimental endotoxemia.

Differential analysis, which covers one-way ANOVA, t test and the related plots has been developed internally, code is available in supplementary material. A p-value of < 0.05 was considered statistically significant after correction [146].

Overall, 174 metabolites were significantly changed between the two groups according to the described method. PCA was utilized to identify the clusters of these significantly changed compounds, as shown in the supplementary material. Described statistical approach identified metabolites that were significantly changed between the control group and the group that was trained in the behavioral intervention. At the baseline of this study, at time point T= minus 1 hour, 19 annotated metabolites were found to be statistically significant. All 19 were decreased in the intervention group when compared to the controls. Table 7 lists significantly changed metabolites and metabolic pathways associated with them.

Table 7: Significantly changed metabolites between intervention and control group at experimental baseline, T= minus 1 hour before LPS administration. P values between groups were calculated for each metabolite using t test.

Annotation	Intervention/Control p value	Intervention/Control log2 ratio	Class
L-Carnitine	0.0001	-0.3686	Alkylamines
Succinic acid semialdehyde	0.0029	-0.4872	Fatty Acids and Conjugates
PC(14:1(9Z)/22:4(7Z,10Z,13Z,16Z))	0.0031	-0.7081	Glycerophospholipids
PS(41:4)	0.0063	-0.5367	Glycerophospholipids
Stenopalustroside A	0.0071	-1.8404	Flavonoids
Glycerophosphoserine (55:0)	0.0087	-0.4692	Glycerophospholipids
Gly-Ser-His	0.0098	0.4250	Peptide
PC(18:2(9Z,12Z)/18:2(9Z,12Z))	0.0166	-0.4393	Glycerophospholipids
Cys-Cys-Gly-Tyr	0.0209	-1.3794	Peptide
PC(20:3(8Z,11Z,14Z)/18:2(9Z,12Z))	0.0259	-0.4077	Glycerophospholipids

1-18:2-2-18:3-monogalactosyldiacylglycerol	0.0271	-0.7988	
3S-hydroxyhexacosanoyl-CoA	0.0293	-0.3437	Fatty Acyls [FA]
Glycerophosphocholine (38:6)	0.0310	-0.4260	Lipids
Xanthine-8-carboxylate	0.0328	-0.5598	Purine Nucleobase
L-Acetylcarnitine	0.0348	-0.5574	Acylated Ester of Carnitine
PC(P-18:1(11Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.0352	-0.3669	Glycerophospholipids
SM(d18:0/16:1(9Z))	0.0388	-0.2667	Sphingolipids
PC(18:2(9Z,12Z)/20:2(11Z,14Z))	0.0453	-0.3977	Glycerophospholipids
[FA amino(8:0)] 3-amino-octanoic acid	0.0473	-0.4346	Amino Fatty Acids
PC(16:0/16:0)	0.0489	-0.4399	Glycerophospholipids

At T= 1 hour, when the intervention group was practicing their learned breathing techniques for 90 minutes already and one hour after LPS administration to both groups, differential analysis found 23 metabolites to be statistically important.
One of the metabolites was lactic acid which was measured and absolutely quantified in our previous reporting [137] and showed the same trend in metabolomics measurement.

This reproducibility indicates that the samples were not affected by storage at -80 °C for 3 years (Figure 15).

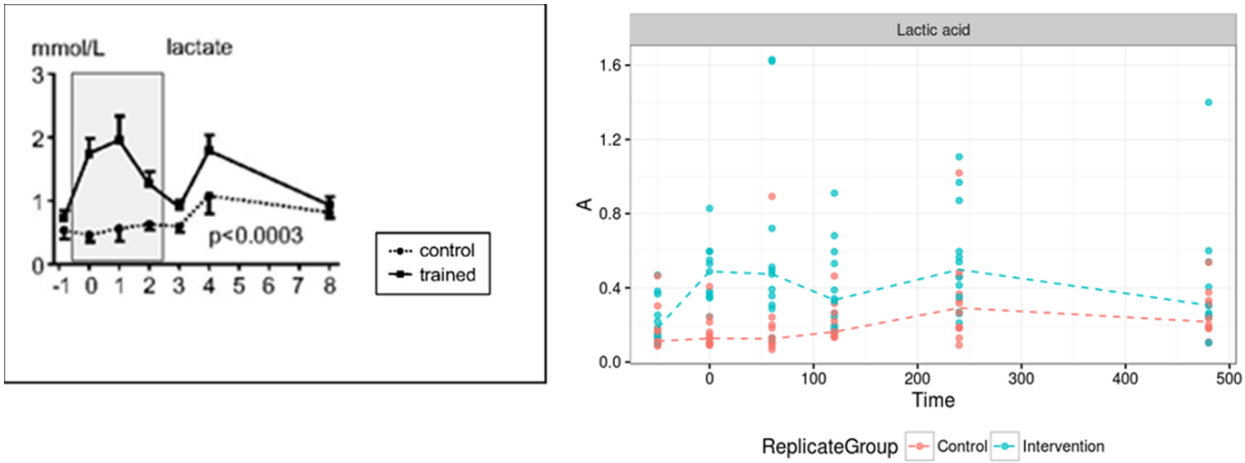


Figure 15: Absolute quantified curve of lactic acid resulting from Kox et al. 2013. (on the left) compared to relative intensities curves of lactic acid (median) resulting from global profiling analysis (on the right).

Two hours after the LPS administration to both groups, at the time point T= 2 hours, 32 metabolites significantly changed, Table 8 lists annotated metabolites.

Table 8: Significantly changed metabolites between intervention and control group at T= 1 hour, 1 hour after LPS administration. P values between groups were calculated for each metabolite using t test

Annotation	Intervention/Control p value	Intervention/Control log2 ratio
L-Carnitine	3.21547E-05	-0.631637312
L-Histidine	0.000456607	-1.595711952
Arginine	0.001848177	-0.718394548
Glutamine	0.002752166	-0.657182892
Succinic acid semialdehyde	0.006775379	-0.742579455
1,6-Naphthalenedisulfonic acid	0.007209301	1.111041353
Deoxyuridine triphosphate	0.007312922	2.351714715
6-Phosphonoglucono-D-lactone	0.008142158	1.511555956
Pyruvic acid	0.012352668	1.482658261
Piperidine	0.013569688	-0.39067304
hexane-6-keto-1,3,4,6-tetracarboxylate	0.014302873	1.690514439
hydroxymethyl-dCDP	0.016124487	-0.557976741
Pyrrolidine	0.021054846	-0.337365318
3-Deoxy-D-glycero-D-galacto-2-nonulosonic acid	0.023427577	1.114202881

N3-fumaramoyl-L-2,3-diaminopropanoate	0.023985261	2.149570973
Lactic acid	0.025882427	1.623274519
N3-(4-methoxyfumaroyl)-L-2,3-diaminopropanoate	0.026178041	1.505049739
LysoPC(20:5(5Z,8Z,11Z,14Z,17Z))	0.035474078	-0.602270254
3'-O-Methyladenosine	0.036879331	-0.312817564
Cer(d18:0/20:0)	0.037376408	-1.032399698
[FA amino(8:0)] 3-amino-octanoic acid	0.039280006	-0.463100982
PC(18:2(9Z,12Z)/20:2(11Z,14Z))	0.043106053	-0.585611723
Deoxycytidine	0.04833029	2.736706038

Changes in metabolite fingerprints of the intervention group were noticeable at every further time point (Tables 9-11).

Table 9: Significantly changed metabolites between intervention and control group at T= 2 hours, 2 hours after LPS administration. P values between groups were calculated for each metabolite using t test

Annotation	Intervention/Control p value	Intervention/Control log2 ratio
Trihydroxy-3,6,7,4'-tetramethoxyflavone	0.000453154	0.633052106
Imidazoleacetic acid riboside	0.00049886	0.503853852
Glucose	0.001778245	0.506498004
Acetic acid	0.003266229	0.581486106
(2Z,4'Z)-2-(5-Methylthio-4-penten-2-ynylidene)-1,6-dioxaspiro[4.4]non-3-ene	0.004792354	0.643716391
3-Deoxy-D-glycero-D-galacto-2-nonulosonic acid	0.00517793	1.093635637
5'-Dehydroadenosine	0.006256427	0.527173281
N-[2-(4-Chloro-phenyl)-acetyl]-N'-(4,7-dimethyl-quinazolin-2-yl)- guanidine	0.006877273	0.805137545
SM(d18:0/16:0)	0.007075819	-1.424118576
L-Acetylcarnitine	0.007959175	-0.838860409
3-O-Caffeoyl-4-O-methylquinic acid	0.00802707	0.762762642
Pyruvic acid	0.008775753	1.001101019
4'-Methoxy-2',3,7-trihydroxyisoflavanone	0.009249669	0.351702995
cis-(homo)2aconitate	0.010613247	0.317167047
PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.011135255	-1.499581685
Pipecolic acid	0.018714721	1.660902676
Prolyl-Gamma-glutamate	0.021398618	0.520703039

PC(o-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.021674782	-0.378903957
PC(18:2(9Z,12Z)/18:2(9Z,12Z))	0.021986358	-0.428642284
N-(2-formyl-3-chlorophenyl)anthranilic acid	0.023161953	-0.12122738
PC(o-16:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.02391054	-0.40226308
2,3-Dimethylbenzofuran	0.024579573	1.191079228
Erucicoyl-EA	0.026110355	-1.856514729
Artonin K	0.026813143	0.458822728
Kaempferol 3-rhamnosyl-(1->2)-[glucosyl-(1->3)-(4'''-p-coumaroylrhamnosyl)-(1->6)-galactoside]	0.028668246	0.244989077
Glyceric acid	0.029843722	0.828594932
Probenecid	0.031060035	0.688706121
6-Phosphonoglucono-D-lactone	0.035558039	0.660784473
2-Hydroxyundecanoate	0.038276913	1.727739069
N-(2-hydroxy-tetracosanoyl)-1-beta-glucosyl-4E,6E-hexadecasphingadienine	0.038858826	-0.395289175
N3-(4-methoxyfumaroyl)-L-2,3-diaminopropanoate	0.042226454	0.748984021
Dimethylarginine	0.043996958	0.226609983
1,6-Naphthalenedisulfonic acid	0.044510103	0.511170137
PE(P-18:1(9Z)/14:0)	0.045007043	0.295286158
[ST (20:4)] cholest-5-en-3beta-yl (15S-hydroperoxy-5Z,8Z,12E,14Z-eicosatetraenoate)	0.046548262	-0.304351104
Proline betaine	0.049002205	1.498820108
SM(d18:0/16:1(9Z))	0.050031937	-0.268109447

Table 10: Significantly changed metabolites between intervention and control group at T= 4 hours, 4 hours after LPS administration. P values between groups were calculated for each metabolite using t test.

Annotation	Intervention/Control p value	Intervention/Control log2 ratio
Trihydroxy-3,6,7,4'-tetramethoxyflavone	0.000453154	0.633052106
Imidazoleacetic acid riboside	0.00049886	0.503853852
Glucose	0.001778245	0.506498004
Acetic acid	0.003266229	0.581486106
(2Z,4'Z)-2-(5-Methylthio-4-penten-2-ynylidene)-1,6-dioxaspiro[4.4]non-3-ene	0.004792354	0.643716391
3-Deoxy-D-glycero-D-galacto-2-nonulosonic acid	0.00517793	1.093635637
5'-Dehydroadenosine	0.006256427	0.527173281

N-[2-(4-Chloro-phenyl)-acetyl]-N'-(4,7-dimethyl-quinazolin-2-yl)- guanidine	0.006877273	0.805137545
SM(d18:0/16:0)	0.007075819	-1.424118576
L-Acetylcarnitine	0.007959175	-0.838860409
3-O-Caffeoyl-4-O-methylquinic acid	0.00802707	0.762762642
Pyruvic acid	0.008775753	1.001101019
4'-Methoxy-2',3,7-trihydroxyisoflavanone	0.009249669	0.351702995
cis-(homo)2aconitate	0.010613247	0.317167047
PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.011135255	-1.499581685
Pipecolic acid	0.018714721	1.660902676
Prolyl-Gamma-glutamate	0.021398618	0.520703039
PC(o-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.021674782	-0.378903957
PC(18:2(9Z,12Z)/18:2(9Z,12Z))	0.021986358	-0.428642284
N-(2-formyl-3-chlorophenyl)anthranilic acid	0.023161953	-0.12122738
PC(o-16:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.02391054	-0.40226308
2,3-Dimethylbenzofuran	0.024579573	1.191079228
Erucicoyl-EA	0.026110355	-1.856514729
Artonin K	0.026813143	0.458822728
Kaempferol 3-rhamnosyl-(1->2)-[glucosyl-(1->3)-(4'''-p-coumaroylrhamnosyl)-(1->6)-galactoside]	0.028668246	0.244989077
Glyceric acid	0.029843722	0.828594932
Probenecid	0.031060035	0.688706121
6-Phosphonoglucono-D-lactone	0.035558039	0.660784473
2-Hydroxyundecanoate	0.038276913	1.727739069
N-(2-hydroxy-tetracosanoyl)-1-beta-glucosyl-4E,6E-hexadecasphingadienine	0.038858826	-0.395289175
N3-(4-methoxyfumaroyl)-L-2,3-diaminopropanoate	0.042226454	0.748984021
Dimethylarginine	0.043996958	0.226609983
1,6-Naphthalenedisulfonic acid	0.044510103	0.511170137
PE(P-18:1(9Z)/14:0)	0.045007043	0.295286158
[ST (20:4)] cholest-5-en-3beta-yl (15S-hydroperoxy-5Z,8Z,12E,14Z-eicosatetraenoate)	0.046548262	-0.304351104
Proline betaine	0.049002205	1.498820108
SM(d18:0/16:1(9Z))	0.050031937	-0.268109447

Table 11: Significantly changed metabolites between intervention and control group at T= 8 hours, 8 hours after LPS administration. P values between groups were calculated for each metabolite using t test.

Annotation	Intervention/Control p value	Intervention/Control log2 ratio
LysoPC(18:2(9Z,12Z))	0.00156078	0.695969939
cis-(homo)2aconitate	0.005863374	-0.225869858
N(6)-Trimethyl-L-lysine	0.006496653	0.457005453
Isocitric acid	0.011816495	0.705278276
PE(P-18:1(9Z)/18:3(6Z,9Z,12Z))	0.012130867	-0.827580957
[ST (20:4)] cholest-5-en-3beta-yl (15S-hydroperoxy-5Z,8Z,12E,14Z-eicosatetraenoate)	0.012247553	-0.431581308
7-methylthioheptanaldoxime	0.014039927	-3.627904943
Ranunculin	0.026072558	0.934292231
Galactaric acid	0.031745087	1.37301928
N-(tetracosanoyl)-sphinganine-1-O-[myo-inositol-1-phosphoryl-6-D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.031994853	-0.312246552
Cer(d18:1/24:1(15Z))	0.032272052	-2.164909433
Glycerophosphocholine	0.033417648	0.510012714
SM(d18:0/16:1(9Z))	0.034190487	-0.297114732
148.0117@0.9983	0.03442392	-1.830026736
N-(2-formyl-3-chlorophenyl)anthranilic acid	0.035216806	-0.149472955
PC(o-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.036427559	-0.358672396
N-(2-hydroxyhexacosanoyl)-4R-hydroxyeicosasphinganine-1-O-[D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.038555695	-0.148511072
Gnidicin	0.038827746	-1.422500552
PC(o-16:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.040539259	-0.396508141
SM(d18:1/18:1(11Z))	0.042167473	-0.599451656
Artonin K	0.044067613	-0.327621346
(2Z,4'Z)-2-(5-Methylthio-4-penten-2-ynylidene)-1,6-dioxaspiro[4.4]non-3-ene	0.046945216	-0.399323188

4.4. Discussion and Conclusion – Part III

In this study two objectives were monitored: metabolite changes in the two groups of healthy individuals induced by breathing techniques and the changes in metabolome in two groups of subjects during experimental endotoxemia. As demonstrated here and fully described in recent review articles and our own reporting, human endotoxemia experiment is providing an insight and understanding of the inflammatory pathways [137, 147]. Subjects that took part in training consisted of meditation, exposure to cold and breathing exercises for 10 days prior to endotoxemia experiment. Baseline of this study showed significant changes in their metabolome when compared to the control group (Table 2), T= minus 1 hour. Metabolite with the highest significance is L-carnitine. It is significantly decreased in the intervention group. About 75% of carnitine is obtained through nutrition in humans [148]. If not obtained from food it is synthesized endogenously from two essential amino acids, lysine and methionine in brain, kidney and liver [149]. Decreased levels of members of the urea cycle (carnitine and acetylcarnitine) as well as amino acids (glutamine, histidine and arginine) indicates that are being utilized as the substrates for energy production.

Furthermore, carnitine levels were decreased in the intervention group through endotoxemia experiment as well. Carnitine levels are reduced in patients suffering Gram-negative sepsis [150] and that prophylactic use of carnitine has been shown to reduce the endothelial damage caused by lipopolysaccharide (LPS) and TNF-α. [151]. Other metabolites that were decreased in the trained group one hour after LPS administration were L-acetylcarnitine, succinic acid semialdehyde, arginine, histidine and glutamine as well as fatty acids and lipids. Furthermore, pyruvic acid and lactic acid were significantly elevated in the trained group. Gene transcription data reported findings of significant increase of lactate production and NAD+ concentration when cells were stimulated with LPS or *C. albicans* [152]. Taken together these findings show that an important mechanism of host defense is the shift from oxidative phosphorylation to aerobic glycolysis, known as the Warburg effect [153]. Inflammation is characterized with metabolic changes in essential amino acids, arginine. Protein breakdown is increased in order to increase arginine availability and via de novo synthesis of arginine is increased in moderate inflammation, while in severe inflammation arginine levels are reduced [154,

155]. Interestingly, levels of essential amino acid arginine were found decreased in the intervention group during the endotoxemia experiment. Furthermore, four hours after the LPS administration to both groups, at T= 4 hours, levels of dimethylarginine were significantly increased in the trained group. Similar results were observed in another study where ratios of arginine, dimethylarginine, citrullin and ornithine were analyzed in septic shock patients. In this study it was observed that the arginine/ornithine ratio was decreased in sepsis, indicating increased arginase activity (an alternative pathway for arginine metabolism) [156]. In addition to decreased arginine levels, histidine is found to be significantly decreased in the trained group as well. These findings are in line with report by Kamisoglu et al., plasma profiling in humans during the endotoxemia experiment. Second cluster of that study displayed significantly increased levels of several amino acids, including proteinogenic amino acids such as histidine and a core member of urea cycle, citrullin, which is known to be an effective arginine precursor [157].These findings were described as a strong indication of the involvement of amino acids in the immediate response to inflammatory insult [158]. Our analysis discovered several metabolites of energy metabolism to be significantly increased in the trained group during the endotoxemia experiment (glucose, isocitric acid, acetic acid, lactic acid, pyruvic acid, glyceric acid and 6-phosphoglucono-D-lactone). Study by Agwunobi et al. reported in their investigation on the the effects of LPS on healthy volunteers. In their reporting LPS induced a fever, tachycardia, and mild arterial hypotension while glucose utilization was increased significantly 120 min after LPS administration but then declined progressively. The reduction in glucose utilization was related to impaired nonoxidative glucose disposal [159]. Also in line with our findings is the study by Wellhoener et al, where intravenous endotoxemia was given to heathy volunteers and it significant increase in levels of lactate, glycerol and pyruvate 60 to 90 minutes after the LPS was administered [160]. In agreement with our data, in studies of human diseases energy-related changes are commonly seen in metabolomics which provide an indicator of disease severity [161]. Interestingly, glucose levels were increased in the trained group when compared to the control group. Additionally, several research groups reported significant increases in concentration of organ failure-related compounds: lactate, 2-hydroxyisovalerate, isobutyrate, creatinine, trimethylamine *N*-oxide, and urea [162, 163].

All together, these data confirm direct implication of LPS to altered energy, lipid and protein metabolism, supporting our findings.

4.5. Acknowledgement – Part III

We greatly appreciate the help of Manuela Hische and Michael Heiser for their support with the analysis, Judith Lenk and Monique Thiele for the technical assistance.

This work was funded by Deutsche Forschungsgemeinschaft (German Research Foundation), Bundesministerium für Bildung und Forschung (BMBF, the Federal Ministery for Education and Research) and we are grateful to the Cusanuswerk (KAAD) for support.

5 Patent for diagnosis of urogenital tract infections: Part IV

Inventors: Dieter Jahn, Nicolas Schauer, Ivana Blaženović

Ref: 856-3 EP

5.1. Summary of the invention

In a first aspect, the present invention relates to a method of diagnosing a urinary tract infection (UTI) in a patient comprising the step of:

determining the level of one or more metabolites in a biological sample from a patient, wherein said one or more metabolites are selected from the group consisting of:

heneicosanoyl-glycero-phospho-(glycerol), thienopyridine, and a metabolite with the sum formula $C_9H_{17}NO$, wherein the O is present in a hydroxy group, the N is present in a tertiary amine group, and the metabolite comprises a -C≡C-H group.

In a second aspect, the present invention relates to a method of determining whether a patient responds to a therapeutic treatment of a urinary tract infection comprising the step of: determining the level of one or more metabolites in a biological sample from a patient, wherein said one or more metabolites are selected from the group consisting of: heneicosanoyl-glycero-phospho-(glycerol), thienopyridine, and a metabolite with the sum formula $C_9H_{17}NO$, wherein the O is present in a hydroxy group, the N is present in a tertiary amine group, and the metabolite comprises a -C≡C-H group.

In a third aspect, the present invention relates to the use of one or more metabolites for diagnosing a urinary tract infection in a patient or for determining whether a patient responds to a therapeutic treatment of a urinary tract infection, wherein said one or more metabolites are selected from the group consisting of: heneicosanoyl-glycero-phospho-(glycerol), thienopyridine, and a metabolite with the sum formula $C_9H_{17}NO$, wherein the

O is present in a hydroxy group, the N is present in a tertiary amine group, and the metabolite comprises a -C≡C-H group.

In a fourth aspect, the present invention relates to a kit comprising means for determining the level of one or more metabolites in a biological sample from a patient, wherein said one or more metabolites are selected from the group consisting of:

heneicosanoyl-glycero-phospho-(glycerol), thienopyridine, and a metabolite with the sum formula $C_9H_{17}NO$, wherein the O is present in a hydroxy group, the N is present in a tertiary amine group, and the metabolite comprises a -C≡C-H group.

This summary of the invention does not necessarily describe all features of the present invention. Other embodiments will become apparent from a review of the ensuing detailed description.

5.2. Acknowledgement – Part IV

We greatly appreciate the help of Judith Lenk and Monique Thiele for the technical assistance and Metabolomic Discoveries GmbH for providing state of the art metabolomics platform.

We are very thankful for support from Prof. Dr. med. Peter Hammerer from Braunschweig Klinikum and the whole staff at Urology department.

Furthermore, we are very thankful to TU Braunschweig for supporting this research.

6 Summary

In the first part of the thesis a metabolomics approach was undertaken for the identification of biomarkers in human urine indicating a urinary tract infection (UTI). For this purpose, the urine of 44 patients suffering from UTI and 48 healthy persons was analyzed for their metabolite composition using LC-MS analysis. For the accurate identification of metabolites indicative of UTI the statistical approach for LC-MS data interpretation had to be optimized. For this purpose, a prognostic model was developed and the relation between the metabolite level and the infection status was calculated by multiple linear regression analysis with the least squares method separately for each metabolite. Utilization of this novel data interpretation approach led to identification of thieno-(2,3-c) pyridine and heneicosanoyl-glycero-3-phospho-(1'-glycerol) as metabolites safely indicating UTIs.

In a second project, software used for putative biomarker annotation, four *in silico* fragmentation tools (MetFragCL, CFM-ID, MAGMa+ and MS-Finder) and their performance was addressed on a Critical Assessment of Small Molecule Identification (CASMI) 2016 dataset. Pure *in silico* algorithms only identified 17% of the compounds correctly. The voting/consensus model developed for this purpose slightly improved the results to 22%. Once the database and MS/MS search was added the algorithm was able to correctly identify 87-93% of the unknown compounds as the first hit, showing the importance of ranking and database queries. These results confirmed that voting/consensus models can be used for real-world applications. Our software will also allow for automatic testing and performance tuning without user interaction.

In a third project, metabolomic profiling was performed on plasma samples obtained in a previously published parallel randomized controlled study by Kox *et.al.*, 2013. Metabolic signatures revealed altered lipid, protein and central metabolism in both intervention and control group as a response to endotoximia and breathing techniques.

Overall, obtained results underscore the increasing power and relevance of metabolomics in biomedical research and diagnosis.

7 Zusammenfassung

Im ersten Teil der Arbeit wurde eine Metabolom Analyse zur Identifikation von Biomarkern in humanem Urin, welche auf eine Harnwegsinfektion (HWI) hinweisen, durchgeführt. Zu diesem Zweck wurden Urinproben von 44 an HWI leidenden und 48 gesunde Personen auf die Komposition ihrer Metabolite mittels LC-MS analysiert. Um eine korrekte Identifikation der auf HWI hindeutenden Metabolite zu gewährleisten musste die statistische Herangehensweise der LC-MS Dateninterpretation optimiert werden. Zu diesem Zweck wurde ein prognostisches Model entwickelt, bei welchem die Beziehung zwischen der Metabolit Konzentration und dem Stand der Infektion mit Hilfe linearer Regressionsanalyse mit der Kleinstquadratmethode individuell für jedes Metabolit ermittelt werden konnte. Die Anwendung dieses neu entwickelten Datenanalysesystems führte zur Identifikation von Thieno-(2,3-c) Pyridin und Heneicosanoyl-glycero-3-phosphp-(1´-glycerol) als Metabolite für die sichere Identifikation einer HWI.

In einem zweiten Projekt wurde eine Software zur Annotation potentieller Biomarker und vier *in silico* Fragmentierungstools (MetFragCL, CFM-ID, MAGMa+ und MS-Finder) nach ihrer Performanz durch die Anwendung auf ein Critical Assessment of Small Molecule Identification (CASMI) 2016 Datenset beurteilt. Die *in silico* Algorithmen alleine haben 17 % der Komponenten korrekt identifiziert. Ein "voting/consensus" Model wurde für diesen Test entwickelt und erzielte eine Verbesserung der Ergebnisse auf 22%. Die zusätzliche Integration einer Datenbank und MS/MS Suche verbesserte die korrekte Identifikation unbekannter Komponenten mittels der Algorithmen auf 87-93% für den ersten Treffer. Zusammenfassend wurde gezeigt, dass das "voting/consensus" Model für reale Anwendungen genutzt werden kann. Zudem erlaubt unsere entwickelte Software sowohl automatisierte Tests als auch eine automatisierte Performanzverbesserung.

In einem dritten Projekt wurde ein metabolisches Profiling an Plasmaprobem durchgeführt, welche im Rahmen einer parallel randomisierten kontrollierten Studie von Kox *et al.* 2013 gesammelt wurden. Diese Analyse des Metaboloms zeigte eine veränderte Komposition der Lipide, Proteine und Metabolite des zentralen Stoffwechsels

sowohl in der Interventions- als auch in der Kontrollgruppe als Reaktion auf Endotoxämie und Atemmethoden.

Zusammenfassend verdeutlichen die gewonnenen Erkenntnisse die zunehmende Bedeutung und Relevanz von Metabolom Analysen in der biomedizinischen Forschung und Diagnose.

8 Outlook

This work investigated the metabolic profiles of patients suffering from urinary tract infections and healthy individuals that underwent human endotoxemia. Furthermore, it was investigated the performance of tools that aid structure elucidation in metabolomics studies. The results raise new questions in terms of data analysis, design of experiments and annotation of MS/MS spectra.

1. Prognostic model developed for metabolomics – based identification of urinary tract infections identified two host-derived compounds indicative of a urinary tract infection: a) Thieno - (2,3-c) pyridine and b) Heneicosanoyl-glycero-3-phospho-(1'-glycerol). However, in order to truly validate the importance of these compounds, a cohort study is needed with as many meta – data as possible. Furthermore, the indicative putative biomarkers need to be synthetized, fragmented and compared against experimental data in order to confirm the true identity of these molecules.

2. Developed voting/consensus model although it improves the results of the four tools tested (CMF – ID, MAGMa+, MetFragCL and MS – Finder), it depends on the results of the underlying tools. Once more and more MS/MS spectra are becoming available and corresponding structural diversity increases, these can be used to train and optimize *in silico* algorithms which will lead to better performance.

3. Metabolomics data obtained from 24 healthy individuals divided into intervention and control group confirm direct implication of LPS to altered energy, lipid and protein metabolism, supporting previous findings. However, sample size needs to be increased in order to avoid model overfitting.

9 References

1. Patti, G.J., O. Yanes, and G. Siuzdak, *Innovation: Metabolomics: the apogee of the omics trilogy.* Nat Rev Mol Cell Biol, 2012. **13**(4): p. 263-9.
2. Zhang, A., et al., *Modern analytical techniques in metabolomics analysis.* Analyst, 2012. **137**(2): p. 293-300.
3. de Rooij, B.M., et al., *Urinary excretion of N-acetyl-S-allyl-L-cysteine upon garlic consumption by human volunteers.* Arch Toxicol, 1996. **70**(10): p. 635-9.
4. Vincent, I.M. and M.P. Barrett, *Metabolomic-based strategies for anti-parasite drug discovery.* J Biomol Screen, 2015. **20**(1): p. 44-55.
5. Schauer, N., et al., *Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement.* Nat Biotechnol, 2006. **24**(4): p. 447-54.
6. Medina, S., et al., *Metabolomics and the diagnosis of human diseases--a guide to the markers and pathophysiological pathways affected.* Curr Med Chem, 2014. **21**(7): p. 823-48.
7. Klupczynska, A., P. Derezinski, and Z.J. Kokot, *Metabolomics in Medical Sciences--Trends, Challenges and Perspectives.* Acta Pol Pharm, 2015. **72**(4): p. 629-41.
8. Fuhrer, T. and N. Zamboni, *High-throughput discovery metabolomics.* Curr Opin Biotechnol, 2015. **31**: p. 73-8.
9. Khamis, M.M., D.J. Adamko, and A. El-Aneed, *Mass spectrometric based approaches in urine metabolomics and biomarker discovery.* Mass Spectrom Rev, 2015.DOI: 10.1002/mas.21455.
10. Theodoridis, G., H.G. Gika, and I.D. Wilson, *Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies.* Mass Spectrom Rev, 2011. **30**(5): p. 884-906.
11. Roberts, L.D., et al., *Targeted metabolomics.* Curr Protoc Mol Biol, 2012. **Chapter 30**: p. Unit 30 2 1-24.
12. Cevallos-Cevallos, J.M. and J.I. Reyes-De-Corcuera, *Metabolomics in food science.* Adv Food Nutr Res, 2012. **67**: p. 1-24.
13. Xi, B., et al., *Statistical analysis and modeling of mass spectrometry-based metabolomics data.* Methods Mol Biol, 2014. **1198**: p. 333-53.
14. Hiller, K., et al., *MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis.* Anal Chem, 2009. **81**(9): p. 3429-39.
15. Cajka, T. and O. Fiehn, *Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics.* Anal Chem, 2016. **88**(1): p. 524-45.

16. West, J.A., et al., *A targeted metabolomics assay for cardiac metabolism and demonstration using a mouse model of dilated cardiomyopathy.* Metabolomics, 2016. **12**: p. 59.

17. Gu, H., et al., *Globally Optimized Targeted Mass Spectrometry: Reliable Metabolomics Analysis with Broad Coverage.* Anal Chem, 2015. **87**(24): p. 12355-62.

18. Zhang, X., et al., *Non-targeted and targeted metabolomics approaches to diagnosing lung cancer and predicting patient prognosis.* Oncotarget, 2016.

19. Hou, W., et al., *A strategy for the targeted metabolomics analysis of 11 gut microbiota-host co-metabolites in rat serum, urine and feces by ultra high performance liquid chromatography-tandem mass spectrometry.* J Chromatogr A, 2016. **1429**: p. 207-17.

20. Klepacki, J., et al., *A high-performance liquid chromatography-tandem mass spectrometry-based targeted metabolomics kidney dysfunction marker panel in human urine.* Clin Chim Acta, 2015. **446**: p. 43-53.

21. Irakli, M.N., V.F. Samanidou, and I.N. Papadoyannis, *Optimization and validation of the reversed-phase high-performance liquid chromatography with fluorescence detection method for the separation of tocopherol and tocotrienol isomers in cereals, employing a novel sorbent material.* J Agric Food Chem, 2012. **60**(9): p. 2076-82.

22. Alonso, A., et al., *AStream: an R package for annotating LC/MS metabolomic data.* Bioinformatics, 2011. **27**(9): p. 1339-40.

23. Kuhl, C., et al., *CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets.* Anal Chem, 2012. **84**(1): p. 283-9.

24. Chong, E.Y., et al., *Local false discovery rate estimation using feature reliability in LC/MS metabolomics data.* Sci Rep, 2015. **5**: p. 17221.

25. Mahieu, N.G., et al., *Defining and Detecting Complex Peak Relationships in Mass Spectral Data: The Mz.unity Algorithm.* Anal Chem, 2016. **88**(18): p. 9037-46.

26. Furey, A., et al., *Ion suppression; a critical review on causes, evaluation, prevention and applications.* Talanta, 2013. **115**: p. 104-22.

27. Johnson, C.H., et al., *Xenobiotic metabolomics: major impact on the metabolome.* Annu Rev Pharmacol Toxicol, 2012. **52**: p. 37-56.

28. Wilson, I.D., et al., *High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies.* J Proteome Res, 2005. **4**(2): p. 591-8.

29. Spagou, K., et al., *HILIC-UPLC-MS for exploratory urinary metabolic profiling in toxicological studies.* Anal Chem, 2011. **83**(1): p. 382-90.

30. Fave, G., et al., *Development and validation of a standardized protocol to monitor human dietary exposure by metabolite fingerprinting of urine samples.* Metabolomics, 2011. **7**(4): p. 469-484.

31. Members, M.S.I.B., et al., *The metabolomics standards initiative.* Nat Biotechnol, 2007. **25**(8): p. 846-8.

32. Karlsson, E.K., D.P. Kwiatkowski, and P.C. Sabeti, *Natural selection and infectious disease in human populations.* Nat Rev Genet, 2014. **15**(6): p. 379-93.

33. Durot, M., P.Y. Bourguignon, and V. Schachter, *Genome-scale models of bacterial metabolism: reconstruction and applications.* FEMS Microbiol Rev, 2009. **33**(1): p. 164-90.

34. Pacchiarotta, T., A.M. Deelder, and O.A. Mayboroda, *Metabolomic investigations of human infections.* Bioanalysis, 2012. **4**(8): p. 919-25.

35. Peng, X., et al., *Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling.* MBio, 2010. **1**(5) DOI: 10.1128/mBio.00206-10.

36. Tang, J., *Microbial metabolomics.* Curr Genomics, 2011. **12**(6): p. 391-403.

37. Patel, J.B., *16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory.* Mol Diagn, 2001. **6**(4): p. 313-21.

38. Woese, C.R. and G.E. Fox, *Phylogenetic structure of the prokaryotic domain: the primary kingdoms.* Proc Natl Acad Sci U S A, 1977. **74**(11): p. 5088-90.

39. Woese, C.R., et al., *A phylogenetic definition of the major eubacterial taxa.* Syst Appl Microbiol, 1985. **6**: p. 143-51.

40. Pace, N.R., *A molecular view of microbial diversity and the biosphere.* Science, 1997. **276**(5313): p. 734-40.

41. Stahl, D.A., et al., *Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences.* Science, 1984. **224**(4647): p. 409-11.

42. Woese, C.R., O. Kandler, and M.L. Wheelis, *Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4576-9.

43. Stamm, W.E. and S.R. Norrby, *Urinary tract infections: disease panorama and challenges.* J Infect Dis, 2001. **183 Suppl 1**: p. S1-4.

44. Hooton, T.M., *Clinical practice. Uncomplicated urinary tract infection.* N Engl J Med, 2012. **366**(11): p. 1028-37.

45. Nielubowicz, G.R. and H.L. Mobley, *Host-pathogen interactions in urinary tract infection.* Nat Rev Urol, 2010. **7**(8): p. 430-41.

46. Lichtenberger, P. and T.M. Hooton, *Complicated urinary tract infections.* Curr Infect Dis Rep, 2008. **10**(6): p. 499-504.

47. Kline, K.A., et al., *Immune activation and suppression by group B streptococcus in a murine model of urinary tract infection.* Infect Immun, 2011. **79**(9): p. 3588-95.

48. Ronald, A., *The etiology of urinary tract infection: traditional and emerging pathogens.* Am J Med, 2002. **113 Suppl 1A**: p. 14S-19S.

49. Kostakioti, M., S.J. Hultgren, and M. Hadjifrangiskou, *Molecular blueprint of uropathogenic Escherichia coli virulence provides clues toward the development of anti-virulence therapeutics.* Virulence, 2012. **3**(7): p. 592-4.

50. Subcommittee on Urinary Tract Infection, S.C.o.Q.I., Management, and K.B. Roberts, *Urinary tract infection: clinical practice guideline for the diagnosis and management of the initial UTI in febrile infants and children 2 to 24 months.* Pediatrics, 2011. **128**(3): p. 595-610.

51. Becknell, B., et al., *The diagnosis, evaluation and treatment of acute and recurrent pediatric urinary tract infections.* Expert Rev Anti Infect Ther, 2015. **13**(1): p. 81-90.

52. Bouatra, S., et al., *The human urine metabolome.* PLoS One, 2013. **8**(9): p. e73076.

53. Chao, M.R., et al., *Urinary nitrite/nitrate ratio measured by isotope-dilution LC-MS/MS as a tool to screen for urinary tract infections.* Free Radic Biol Med, 2016. **93**: p. 77-83.

54. Hu, C.W., et al., *Elevated urinary levels of carcinogenic N-nitrosamines in patients with urinary tract infections measured by isotope dilution online SPE LC-MS/MS.* J Hazard Mater, 2016. **310**: p. 207-16.

55. Gupta, A., et al., *(1)H NMR spectroscopy in the diagnosis of Pseudomonas aeruginosa-induced urinary tract infection.* NMR Biomed, 2005. **18**(5): p. 293-9.

56. Lam, C.W., et al., *NMR-based metabolomic urinalysis: a rapid screening test for urinary tract infection.* Clin Chim Acta, 2014. **436**: p. 217-23.

57. Peng, B., et al., *Exogenous alanine and/or glucose plus kanamycin kills antibiotic-resistant bacteria.* Cell Metab, 2015. **21**(2): p. 249-61.

58. Su, Y.B., et al., *Fructose restores susceptibility of multidrug-resistant Edwardsiella tarda to kanamycin.* J Proteome Res, 2015. **14**(3): p. 1612-20.

59. Vonberg, R.P., et al., *[Costs due to urinary tract infections in Germany. An estimation based on the data from the German National Nosocomial Infections Surveillance System]*. Urologe A, 2008. **47**(1): p. 54-8.

60. Griebling, T.L., *Urologic diseases in america project: trends in resource use for urinary tract infections in men.* J Urol, 2005. **173**(4): p. 1288-94.

61. Ikaheimo, R., et al., *Recurrence of urinary tract infection in a primary care setting: analysis of a 1-year follow-up of 179 women.* Clin Infect Dis, 1996. **22**(1): p. 91-9.

62. Foxman, B., *Recurring urinary tract infection: incidence and risk factors.* Am J Public Health, 1990. **80**(3): p. 331-3.

63. Foxman, B. and J.W. Chi, *Health behavior and urinary tract infection in college-aged women.* J Clin Epidemiol, 1990. **43**(4): p. 329-37.

64. Foxman, B., *The epidemiology of urinary tract infection.* Nat Rev Urol, 2010. **7**(12): p. 653-60.

65. Dobrindt, U., *[Virulence factors of uropathogens]*. Urologe A, 2010. **49**(5): p. 598-605.

66. Johnson, J.R., *Virulence factors in Escherichia coli urinary tract infection.* Clin Microbiol Rev, 1991. **4**(1): p. 80-128.

67. Rasmussen, M., *Aerococci and aerococcal infections.* J Infect, 2013. **66**(6): p. 467-74.

68. Sohail, M., et al., *Characteristics and Antibiotic Resistance of Urinary Tract Pathogens Isolated From Punjab, Pakistan.* Jundishapur J Microbiol, 2015. **8**(7): p. e19272.

69. Wilson, M.L. and L. Gaido, *Laboratory diagnosis of urinary tract infections in adult patients.* Clin Infect Dis, 2004. **38**(8): p. 1150-8.

70. Franz, M. and W.H. Horl, *Common errors in diagnosis and management of urinary tract infection. II: clinical management.* Nephrol Dial Transplant, 1999. **14**(11): p. 2754-62.

71. Mishra, B., et al., *Symptom-based diagnosis of urinary tract infection in women: are we over-prescribing antibiotics?* Int J Clin Pract, 2012. **66**(5): p. 493-8.

72. Leushuis, E., et al., *Prediction models in reproductive medicine: a critical appraisal.* Hum Reprod Update, 2009. **15**(5): p. 537-52.

73. Agrawal, P., et al., *Role of blood C - reactive protein levels in upper urinary tract infection and lower urinary tract infection in adult patients (>16 years).* J Assoc Physicians India, 2013. **61**(7): p. 462-3.

74. Vorholter, F.J., et al., *Genome Sequence of the Urethral Catheter Isolate Pseudomonas aeruginosa MH19.* Genome Announc, 2015. **3**(2).

75. Lassek, C., et al., *A metaproteomics approach to elucidate host and pathogen protein expression during catheter-associated urinary tract infections (CAUTIs).* Mol Cell Proteomics, 2015. **14**(4): p. 989-1008.

76. Lam, C.W., et al., *Quantitative metabolomics of urine for rapid etiological diagnosis of urinary tract infection: evaluation of a microbial-mammalian co-metabolite as a diagnostic biomarker.* Clin Chim Acta, 2015. **438**: p. 24-8.

77. Lam, C.W., et al., *NMR-based urinalysis for rapid diagnosis of beta-ureidopropionase deficiency in a patient with Dravet syndrome.* Clin Chim Acta, 2015. **440**: p. 201-4.

78. Lisec, J., et al., *Corrigendum: Gas chromatography mass spectrometry-based metabolite profiling in plants.* Nat Protoc, 2015. **10**(9): p. 1457.

79. Cristianini, N. and B. Scholkopf, *Support vector machines and kernel methods - The new generation of learning machines.* Ai Magazine, 2002. **23**(3): p. 31-41.

80. Sahiner, B., et al., *Design of a high-sensitivity classifier based on a genetic algorithm: application to computer-aided diagnosis.* Phys Med Biol, 1998. **43**(10): p. 2853-71.

81. Crisci, C., B. Ghattas, and G. Perera, *A review of supervised machine learning algorithms and their applications to ecological data.* Ecological Modelling, 2012. **240**: p. 113-122.

82. Sajda, P., *Machine learning for detection and diagnosis of disease.* Annu Rev Biomed Eng, 2006. **8**: p. 537-65.

83. Hastie, T., R. Tibshirani, and J. Friedman, *Note on "Comparison of model selection for regression" by Vladimir Cherkassky and Yunqian Ma.* Neural Comput, 2003. **15**(7): p. 1477-80.

84. Cherkassky, V. and Y. Ma, *Another look at statistical learning theory and regularization.* Neural Netw, 2009. **22**(7): p. 958-69.

85. Phan, J.H., C.F. Quo, and M.D. Wang, *Functional genomics and proteomics in the clinical neurosciences: data mining and bioinformatics.* Prog Brain Res, 2006. **158**: p. 83-108.

86. Toronen, P., et al., *Analysis of gene expression data using self-organizing maps.* FEBS Lett, 1999. **451**(2): p. 142-6.

87. Haybittle, J.L., et al., *A prognostic index in primary breast cancer.* Br J Cancer, 1982. **45**(3): p. 361-6.

88. Melander, O., et al., *Novel and conventional biomarkers for prediction of incident cardiovascular events in the community.* JAMA, 2009. **302**(1): p. 49-57.

89. Peng, J., et al., *Development of a universal metabolome-standard method for long-term LC-MS metabolome profiling and its application for bladder cancer urine-metabolite-biomarker discovery.* Anal Chem, 2014. **86**(13): p. 6540-7.

90. Baker, M., *Metabolomics: from small molecules to big ideas.* Nature Methods, 2011. **8**(2): p. 117-121.

91. Allen, F., et al., *CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra.* Nucleic Acids Research, 2014. **42**(W1): p. W94-W99.

92. Ruttkies, C., et al., *MetFrag relaunched: incorporating strategies beyond in silico fragmentation.* Journal of Cheminformatics, 2016. **8**.

93. de Toro-Peinado, I., et al., *[Microbiological diagnosis of urinary tract infections]*. Enferm Infecc Microbiol Clin, 2015. **33 Suppl 2**: p. 34-9.

94. Chakravorty, S., et al., *A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria.* J Microbiol Methods, 2007. **69**(2): p. 330-9.

95. Schwieger, F. and C.C. Tebbe, *A new approach to utilize PCR-single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis.* Appl Environ Microbiol, 1998. **64**(12): p. 4870-6.

96. Dohrmann, A.B. and C.C. Tebbe, *Effect of elevated tropospheric ozone on the structure of bacterial communities inhabiting the rhizosphere of herbaceous plants native to Germany.* Appl Environ Microbiol, 2005. **71**(12): p. 7750-8.

97. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.

98. Brand, B., et al., *Temperament type specific metabolite profiles of the prefrontal cortex and serum in cattle.* PLoS One, 2015. **10**(4): p. e0125044.

99. van de Beek, M.C., et al., *C26:0-Carnitine Is a New Biomarker for X-Linked Adrenoleukodystrophy in Mice and Man.* PLoS One, 2016. **11**(4): p. e0154597.

100. Technologies, A. 2011.

101. Eduati, F., et al., *A Boolean approach to linear prediction for signaling network modeling.* PLoS One, 2010. **5**(9).

102. Aickin, M. and H. Gensler, *Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods.* Am J Public Health, 1996. **86**(5): p. 726-8.

103. Paul Gregor, N.H., Juraj Koppel, Regina Zhuk, *Pharmaceutical compositions comprising thieno[2,3-c]pyridine derivatives and use thereof* 2004.

104. Martin, R.S. and D.M. Small, *Physicochemical characterization of the urinary lipid from humans with nephrotic syndrome.* J Lab Clin Med, 1984. **103**(5): p. 798-810.

105. Saunders, D.A., et al., *Measurement of total phospholipids in urine of patients treated with gentamicin.* Br J Clin Pharmacol, 1997. **43**(4): p. 435-40.

106. Hirayama, A., et al., *Metabolic profiling reveals new serum biomarkers for differentiating diabetic nephropathy.* Anal Bioanal Chem, 2012. **404**(10): p. 3101-9.

107. Ruttkies, C., M. Gerlich, and S. Neumann, *Tackling CASMI 2012: Solutions from MetFrag and MetFusion.* Metabolites, 2013. **3**(3): p. 623-36.

108. Robin, X., et al., *pROC: an open-source package for R and S plus to analyze and compare ROC curves.* Bmc Bioinformatics, 2011. **12**.

109. Pepe, M.S., T. Cai, and G. Longton, *Combining predictors for classification using the area under the receiver operating characteristic curve.* Biometrics, 2006. **62**(1): p. 221-9.

110. Zhao, L.P., et al., *An Object-Oriented Regression for Building Disease Predictive Models with Multiallelic HLA Genes.* Genet Epidemiol, 2016. **40**(4): p. 315-32.

111. Hufsky, F. and S. Böcker, *Mining molecular structure databases: Identification of small molecules based on fragmentation mass spectrometry data.* Mass spectrometry reviews, 2016.

112. Allard, P.-M., et al., *Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication.* Analytical chemistry, 2016. **88**(6): p. 3317-3323.

113. Ridder, L., et al., *Automatic Chemical Structure Annotation of an LC–MS n Based Metabolic Profile from Green Tea.* Analytical chemistry, 2013. **85**(12): p. 6033-6040.

114. Ridder, L., J.J. van der Hooft, and S. Verhoeven, *Automatic compound annotation from mass spectrometry data using MAGMa.* Mass Spectrometry, 2014. **3**(Spec Iss 2).

115. Verdegem, D., et al., *Improved metabolite identification with MIDAS and MAGMa through MS/MS spectral dataset-driven parameter optimization.* Metabolomics, 2016. **12**(6): p. 1-16.

116. Meringer, M., et al., *MS/MS data improves automated determination of molecular formulas by mass spectrometry.* MATCH Commun. Math. Comput. Chem, 2011. **65**: p. 259-290.

117. Dührkop, K., et al., *Searching molecular structure databases with tandem mass spectra using CSI: FingerID.* Proceedings of the National Academy of Sciences, 2015. **112**(41): p. 12580-12585.

118. Allen, F., R. Greiner, and D. Wishart, *Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification.* Metabolomics, 2015. **11**(1): p. 98-110.

119. Horai, H., et al., *MassBank: a public repository for sharing mass spectral data for life sciences.* Journal of Mass Spectrometry, 2010. **45**(7): p. 703-714.

120. Schymanski, E.L., et al., *Solving CASMI 2013 with MetFrag, MetFusion and MOLGEN-MS/MS.* Mass Spectrometry, 2014. **3**(Spec Iss 2).

121. Tsugawa, H., et al., *Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software.* Analytical Chemistry, 2016.

122. Schymanski, E.L. and S. Neumann, *The Critical Assessment of Small Molecule Identification (CASMI): Challenges and Solutions.* Metabolites, 2013. **3**(3): p. 517-38.

123. Ridder, L., et al., *Substructure‐based annotation of high‐resolution multistage MSn spectral trees.* Rapid Communications in Mass Spectrometry, 2012. **26**(20): p. 2461-2471.

124. Grimme, S., *Towards first principles calculation of electron impact mass spectra of molecules.* Angewandte Chemie International Edition, 2013. **52**(24): p. 6306-6312.

125. Huckele, S. and T. Track, *Risk management of emerging compounds and pathogens in the water cycle (RiSKWa).* Environmental Sciences Europe, 2013. **25**(1): p. 1-4.

126. Jeffryes, J.G., et al., *MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics.* J Cheminform, 2015. **7**: p. 44.

127. Linstrom, P.J. and W.G. Mallard, *The NIST Chemistry WebBook: A chemical data resource on the internet.* Journal of Chemical and Engineering Data, 2001. **46**(5): p. 1059-1063.

128. Williams, A. and V. Tkachenko, *The Royal Society of Chemistry and the delivery of chemistry data repositories for the community.* J Comput Aided Mol Des, 2014. **28**(10): p. 1023-30.

129. Oberacher, H., *Applying Tandem Mass Spectral Libraries for Solving the Critical Assessment of Small Molecule Identification (CASMI) LC/MS Challenge 2012.* Metabolites, 2013. **3**(2): p. 312-324.

130. Newsome, A.G. and D. Nikolic, *CASMI 2013: Identification of small molecules by tandem mass spectrometry combined with database and literature mining.* Mass Spectrometry, 2014. **3**(Spec Iss 2).

131. Brouard, C., et al., *Fast metabolite identification with Input Output Kernel Regression.* Bioinformatics, 2016. **32**(12): p. i28-i36.

132. Demarque, D.P., et al., *Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products.* Natural product reports, 2016. **33**(3): p. 432-455.

133. Bravo, E.L., *Metabolic factors and the sympathetic nervous system.* Am J Hypertens, 1989. **2**(12 Pt 2): p. 339S-344S.

134. Neugebauer, S., et al., *Metabolite Profiles in Sepsis: Developing Prognostic Tools Based on the Type of Infection.* Crit Care Med, 2016. **44**(9): p. 1649-62.

135. Schwertner, H.A. and S.B. Kong, *Determination of modafinil in plasma and urine by reversed phase high-performance liquid-chromatography.* J Pharm Biomed Anal, 2005. **37**(3): p. 475-9.

136. Mirsaeidi, M., et al., *Plasma metabolomic profile in fibrosing pulmonary sarcoidosis.* Sarcoidosis Vasc Diffuse Lung Dis, 2016. **33**(1): p. 29-38.

137. Kox, M., et al., *Voluntary activation of the sympathetic nervous system and attenuation of the innate immune response in humans.* Proc Natl Acad Sci U S A, 2014. **111**(20): p. 7379-84.

138. Smith, C.A., et al., *XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification.* Anal Chem, 2006. **78**(3): p. 779-87.

139. Libiseller, G., et al., *IPO: a tool for automated optimization of XCMS parameters.* BMC Bioinformatics, 2015. **16**: p. 118.

140. Team, R.C., *R: A Language and Environment for Statistical Computing*. 2016.

141. Creek, D.J., et al., *IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data.* Bioinformatics, 2012. **28**(7): p. 1048-9.

142. Yamada, T., et al., *iPath2.0: interactive pathway explorer.* Nucleic Acids Res, 2011. **39**(Web Server issue): p. W412-5.

143. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.

144. Wishart, D.S., et al., *HMDB: the Human Metabolome Database.* Nucleic Acids Res, 2007. **35**(Database issue): p. D521-6.

145. Kind, T. and O. Fiehn, *Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry.* BMC Bioinformatics, 2007. **8**: p. 105.

146. Holm, S., *A Simple Sequentially Rejective Multiple Test Procedure.* Scandinavian Journal of Statistics, 1979. **6**(2): p. 65-70.

147. Fullerton, J.N., et al., *Intravenous Endotoxin Challenge in Healthy Humans: An Experimental Platform to Investigate and Modulate Systemic Inflammation.* J Vis Exp, 2016(111).

148. Pons, R. and D.C. De Vivo, *Primary and secondary carnitine deficiency syndromes.* J Child Neurol, 1995. **10 Suppl 2**: p. S8-24.

149. Cave, M.C., et al., *Obesity, inflammation, and the potential application of pharmaconutrition.* Nutr Clin Pract, 2008. **23**(1): p. 16-34.

150. Famularo, G. and C. De Simone, *A new era for carnitine?* Immunol Today, 1995. **16**(5): p. 211-3.

151. Penn, D., et al., *Carnitine deprivation adversely affects cardiovascular response to bacterial endotoxin (LPS) in the anesthetized neonatal pig.* Shock, 1998. **10**(5): p. 377-82.

152. Cheng, S.C., et al., *Broad defects in the energy metabolism of leukocytes underlie immunoparalysis in sepsis.* Nat Immunol, 2016. **17**(4): p. 406-13.

153. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* Science, 2009. **324**(5930): p. 1029-33.

154. Bruins, M.J., et al., *In vivo measurement of nitric oxide production in porcine gut, liver and muscle during hyperdynamic endotoxaemia.* Br J Pharmacol, 2002. **137**(8): p. 1225-36.

155. Luiking, Y.C., et al., *Arginine de novo and nitric oxide production in disease states.* Am J Physiol Endocrinol Metab, 2012. **303**(10): p. E1177-89.

156. Weiss, S.L., et al., *Evaluation of asymmetric dimethylarginine, arginine, and carnitine metabolism in pediatric sepsis.* Pediatr Crit Care Med, 2012. **13**(4): p. e210-8.

157. Urschel, K.L., et al., *Citrulline is an effective arginine precursor in enterally fed neonatal piglets.* J Nutr, 2006. **136**(7): p. 1806-13.

158. Kamisoglu, K., et al., *Temporal metabolic profiling of plasma during endotoxemia in humans.* Shock, 2013. **40**(6): p. 519-26.

159. Agwunobi, A.O., et al., *Insulin resistance and substrate utilization in human endotoxemia.* J Clin Endocrinol Metab, 2000. **85**(10): p. 3770-8.

160. Wellhoener, P., et al., *Metabolic alterations in adipose tissue during the early phase of experimental endotoxemia in humans.* Horm Metab Res, 2011. **43**(11): p. 754-9.

161. Mickiewicz, B., et al., *Metabolic profiling of serum samples by 1H nuclear magnetic resonance spectroscopy as a potential diagnostic approach for septic shock.* Crit Care Med, 2014. **42**(5): p. 1140-9.

162. Clausen, M.R., P.B. Mortensen, and F. Bendtsen, *Serum levels of short-chain fatty acids in cirrhosis and hepatic coma.* Hepatology, 1991. **14**(6): p. 1040-5.

163. Landaas, S. and C. Jakobs, *The occurrence of 2-hydroxyisovaleric acid in patients with lactic acidosis and ketoacidosis.* Clin Chim Acta, 1977. **78**(3): p. 489-93.

Figures used in Introduction in this dissertation were adapted with permission obtained through Copyright Clearance Center's RightsLink service in October 2016.

Figure 1: Central dogma of the "omics" studies. Adapted from Gorrochategui *et al., Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow*, TrAC Trends in Analytical Chemistry, License Number: 3966410173327

Figure 2: Analysis workflow in untargeted metabolomic studies. This figure shows the different steps of the metabolomic analysis pipeline. Adapted from Patt *et al., Metabolomics: the apogee of the omics trilogy*, Nature Reviews Molecular Cell Biology, License number: 3966410848412.

Figure 3: Universal phylogenetic tree based on the 16S rRNA gene sequence comparisons. Adapted from Woese *et al., A phylogenetic definition of the major eubacterial taxa.,* Syst Appl Microbiol.

Figure 4: Epidemiology of urinary tract infections. Adapted from Ana L. Flores-Mireles *et al., Urinary tract infections: epidemiology, mechanisms of infection and treatment options*, Nature Reviews Microbiology, License number: 3966540645314

Appendices

Figures

Figure 1: Central dogma of the "omics" studies (adapted from Gorrochategui *et al*., 2016). ..... - 9 -

Figure 2: Analysis workflow in untargeted metabolomic studies. This figure shows the different steps of the metabolomic analysis pipeline (adapted from Patt *et al.*, 2012).- 11 -

Figure 3: Universal phylogenetic tree based on the 16S rRNA gene sequence comparisons (adapted from Woese *et al.,* 1985). ..... - 16 -

Figure 4: Epidemiology of urinary tract infections (adapted from Ana L. Flores-Mireles *et al,* 2015). ..... - 18 -

Figure 5: General workflow ..... - 25 -

Figure 6: Scores plot of the Principal Component Analysis (PCA) prognostic regression model built with a mass spectral data derived from the from urine training set, A, and the validation set, B. colored by infectious status. Red dots indicate samples from infected patients and green dot samples from the healthy subjects. The first principal component of the training set, A, accounted for 43% while the second principal component accounted for 31% overall variability. ..... - 33 -

Figure 7: The first two principal components of the validation set, B, explain 76% of overall variability. ..... - 34 -

Figure 8: Box plots of the normalized peak intensities of representative putative biomarkers. (A) Thieno-(2,3-c) pyridine, (B) Heneicosanoyl-glycero-3-phospho-(1'-glycerol), (C) $C_9H_{17}NO$..... - 36 -

Figure 9: Receiver operating characteristic (ROC) curves of the putative biomarkers for Thieno-(2,3-c) pyridine (A) in the training set and (B) in the validation set, for Heneicosanoyl-glycero-3-phospho-(1'-glycerol) (C) in the training set and (D) in the validation set and for $C_9H_{17}NO$ (E) in training set and (F) in the validation set ..... - 37 -

Figure 10: Receiver operating characteristic (ROC) curves of the combinatorial putative biomarkers (A) in the training set and (B) in the validation set..... - 38 -

Figure 11: Comparison of all *in silico* fragmentation tools and a voting consensus model on a training data set consisted of 312 MS/MS spectra obtained in both positive and negative ionization mode ..... - 51 -

Figure 12: In silico fragmentation can be used to identify unknown MS/MS spectra by matching theoretical fragments to experimental MS/MS spectra and score all fragmented compounds. ..... - 52 -

Figure 13: Principal component analysis of the molecular descriptor space from the training and validation set. Individual outliers show compounds only found in a specific data set. Overlapping dots show very similar compounds. ..... - 61 -

Figure 14: Heat map describing the highest differences between the trained individuals and control group from T= - 1 hour up to T= 8 hours, during experimental endotoxemia. ..... - 83 -

Figure 15: Absolute quantified curve of lactic acid resulting from Kox et al. 2013. (on the left) compared to relative intensities curves of lactic acid (median) resulting from global profiling analysis (on the right). ..... - 86 -

Tables

Table 1: Overview of all identified bacterial genera found in the investigated urine samples when analyzed with PCR-SSCP method and *in vitro* cultivation. In 21 urine samples more than one bacterial species was found. ..... - 30 -

Table 2: Characteristics of the putative biomarkers for urinary tract infection detected in LC-MS chromatograms with the developed prognostic model ..... - 39 -

Table 3: Results of the *in silico* fragmentation performance for four publicly available software tools: MetFragCL, CFM-ID, MAGMa+ and MS-Finder. DB designates priority ranking by structure importance and MS/MS designates MS/MS library search. The 312 MS/MS spectra were the CASMI 2016 training data and the performance of each tool was compared to the voting/consensus model. ..... - 56 -

Table 4: Results of *in silico* performance of MetFragCL, MS-Finder, CFM-ID and MAGMa+ using 208 MS/MS spectra from the 2016 CASMI contest (validation/challenge set). DB designates structure ranking importance and MS/MS library search. ..... - 58 -

Table 5: Subject demographic characteristics. Parameters were measured during screening visit. BMI, body mass index; HR, heart rate; MAP, mean arterial blood pressure. Data are presented as median (range) ..... - 78 -

Table 6: Parameter settings used for raw data processing ..... - 81 -

Table 7: Significantly changed metabolites between intervention and control group at experimental baseline, T= minus 1 hour before LPS administration. P values between groups were calculated for each metabolite using t test. ..... - 84 -

Table 8: Significantly changed metabolites between intervention and control group at T= 1 hour, 1 hour after LPS administration. P values between groups were calculated for each metabolite using t test ..... - 86 -

Table 9: Significantly changed metabolites between intervention and control group at T= 2 hours, 2 hours after LPS administration. P values between groups were calculated for each metabolite using t test ..... - 87 -

Table 10: Significantly changed metabolites between intervention and control group at T= 4 hours, 4 hours after LPS administration. P values between groups were calculated for each metabolite using t test. ..... - 88 -

Table 11: Significantly changed metabolites between intervention and control group at T= 8 hours, 8 hours after LPS administration. P values between groups were calculated for each metabolite using t test. ..... - 90 -

Supplemental tables for Part III

Table A1 Metabolome changes in the control group at time point 0 hours (before LPS administration) compared to the baseline of this study, - 50 minutes. 1-way ANOVA results in the global.p values, and t-tests are the local / pairwise comparisons (local p).

Metabolite	global.p	global.p.adj	Control 0/Control -50: local.p	Control 0/Control -50: ratio
Propionylcarnitine	0.12	1	0.97	0.01
PC(14:1(9Z)/22:4(7Z,10Z,13Z,16Z))	0.04	1	0.08	-0.40
Kaempferol 3-rhamnosyl-(1->2)-[glucosyl-(1->3)-(4'''-p-coumaroylrhamnosyl)-(1->6)-galactoside]	0.03	1	0.96	0
Phosphocreatinine	0.45	1	0.44	0.17
L-Histidine	0.65	1	0.05	0.77
Hymenocardine	0.34	1	0.44	0.24
Chinomethionat	0.49	1	0.06	-0.84
Glycerophosphoinositol (41:6)	0.17	1	0.08	-0.85
N(6)-Trimethyl-L-lysine	0.01	1	0.15	-0.28
5'-Dehydroadenosine	0.02	1	0.34	-0.14
PE(P-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.36	1	0.19	-0.47
Moracin	0.12	1	0.14	-0.54
N-desmethylclarithromycin	0.38	1	0.57	0.06
Glutamine	0.01	1	0.73	0.05
CDP-Glycerophospholipid(40:6)	0.11	1	0.14	-0.31
PC(P-18:0/15:0)	0.21	1	0.91	-0.11
[FA amino(8:0)] 3-amino-octanoic acid	0	0.38	0.12	-0.29
PI(O-20:0/14:1(9Z))	0.30	1	0.72	-0.09
Urea	0.11	1	0.39	-0.12
Triradylglycero (60:5)	0.23	1	0.72	-0.10
Prolyl-Gamma-glutamate	0.07	1	0.03	-0.30
[PR trimethyl,methyl,methyl(5:0/15:0)] 2,6,14-trimethyl-6,7-epoxy-10-methylene-9-(3-methylpent-4-enyl)-pentadec-2-ene	0.09	1	0.49	-0.28
1-18:2-2-18:3-monogalactosyldiacylglycerol	0.41	1	0.09	-0.53
N(alpha)-t-Butoxycarbonyl-L-leucine	0.24	1	0.20	-0.84
Bruceoside A	0.01	1	0.15	-0.67
Phosphatidyl glycerophosphate (O-42:2)	0.49	1	0.32	-0.29
PS(41:2)	0.11	1	0.68	-0.10
PC(O-16:0/18:2(9Z,12Z))	0.10	1	0.31	-0.19
N-(2-hydroxy-tetracosanoyl)-1-beta-glucosyl-4E,6E-hexadecasphingadienine	0.54	1	0.06	-0.29
LysoPC(18:2(9Z,12Z))	0	0.01	0.15	-0.37
L-Carnitine	0.01	1	0.05	-0.18
PC(o-18:0/20:4(8Z,11Z,14Z,17Z))	0.35	1	0.33	-0.14
Epimedin K	0.52	1	0.87	-0.01
Galactaric acid	0.32	1	0.25	-1.11
Iodipamide	0.73	1	0.66	-0.20

PE(P-18:1(9Z)/18:3(6Z,9Z,12Z))	0.27	1	0.31	-0.30
Ranunculin	0	0.12	0.28	-0.78
Spirolide B	0.60	1	0.13	-0.20
CDP-Glycerophospholipid (18:1(9Z)/18:2(9Z,12Z))	0.32	1	0.73	-0.10
PS(39:4)	0.75	1	0.76	-0.09
3-Amino-1,2,4-triazole	0.52	1	0.93	-0.04
Di-2-thienyl disulfide	0.19	1	0.14	-0.33
[6]-Gingerdiol 3,5-diacetate	0.74	1	0.72	0.19
PC(o-16:1(9Z)/18:2(9Z,12Z))	0.24	1	0.09	-0.26
Glycerophosphocholine (40:6)	0.15	1	0.64	-0.12
PI(O-36:2)	0.25	1	0.50	-0.22
CDP-Glycerophospholipid (38:4)	0.25	1	0.88	-0.04
PC(18:1(9Z)/18:1(9Z))	0.09	1	0.48	-0.12
PE(24:1(15Z)/15:0)	0.17	1	0.43	-0.13
PC(o-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.45	1	0.07	-0.23
Prostaglandin F1a	0.31	1	0.02	1.23
Succinic acid semialdehyde	0.01	1	0.12	-0.27
3-O-Caffeoyl-4-O-methylquinic acid	0.06	1	0.73	-0.08
PC(o-18:2(9Z,12Z)/18:2(9Z,12Z))	0.23	1	0.10	-0.23
PC(P-18:1(11Z)/22:5(7Z,10Z,13Z,16Z,19Z))	0.57	1	0.51	-0.12
Glucose	0	0.47	0.51	-0.08
Oxolucidine B	0	0.02	0.07	-0.44
Beta-1,4-D-Mannosylchitobiosyldiphosphodolichol	0.19	1	0.30	0.12
PE-NMe(18:1(9Z)/18:1(9Z))	0.36	1	0.41	-0.16
Artonin K	0	0	0.07	-0.22
5-Ethoxy-4,5-dihydro-2(3H)furanone	0	0.65	0	-0.39
Didodecyl thiobispropanoate	0.38	1	0.80	-0.13
Erucicoyl-EA	0.75	1	0.40	0.48
PC(16:0/16:0)	0.57	1	0.36	-0.20
Xanthine-8-carboxylate	0.10	1	0.73	0.08
N-(tetracosanoyl)-sphinganine-1-O-[myo-inositol-1-phosphoryl-6-D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.56	1	0.50	0.07
PC(o-16:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.48	1	0.20	-0.19
hexane-6-keto-1,3,4,6-tetracarboxylate	0.25	1	0.75	0.24
Ganglioside GM3 (d18:0/20:0)	0.50	1	0.08	-0.18
Epoxymurin B	0.77	1	0.29	-0.58
carboxynorspermidine	0.59	1	0.58	-0.17
PE(15:0/22:0)	0.33	1	0.42	-0.13
Glycerophosphoserine (55:0)	0.39	1	0.23	-0.17
PS(37:2)	0.34	1	0.57	-0.18
DG(43:5)	0.63	1	0.57	-0.11
3S-hydroxyhexacosanoyl-CoA	0.56	1	0.36	-0.16
Trihexosylceramide (d18:1/24:0)	0.84	1	0.59	0.06
PS(21:0/0:0)	0	0.27	0.22	-0.43
PE(P-18:1(9Z)/14:0)	0.32	1	0.94	-0.01

Creatine	0.72	1	0.32	-0.35
L-Acetylcarnitine	0.01	1	0.30	-0.28
PC(o-16:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.63	1	0.14	-0.24
PE(18:4(6Z,9Z,12Z,15Z)/20:2(11Z,14Z))	0.75	1	0.55	0.26
Tartronate semialdehyde	0.37	1	0.84	-0.11
SM(d34:2)	0.65	1	0.94	-0.01
Sedoheptulose	0.69	1	0.21	-0.23
PE(22:5(7Z,10Z,13Z,16Z,19Z)/16:0)	0.88	1	0.78	-0.13
LysoPE(20:0/0:0)	0.04	1	0.79	-0.19
Diphenylcarbazide	0.65	1	0.58	-0.36
PC(P-18:1(11Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.61	1	0.47	-0.15
CDP-DG(16:0/20:4(5Z,8Z,11Z,14Z))	0.14	1	0.28	-0.16
PE-NMe(16:0/18:1(9Z))	0.54	1	0.71	-0.22
PC(14:1(9Z)/20:2(11Z,14Z))	0.17	1	0.35	-0.19
L-Cyclo(alanylglycyl)	0.77	1	0.77	-0.22
PG(O-36:1)	0.26	1	0.38	-0.15
CDP-Glycerophospholipid (38:6)	0.67	1	0.55	-0.13
D-Sedoheptulose 7-phosphate	0.87	1	0.78	-0.07
PS(41:4)	0.44	1	0.35	-0.20
PE(18:4(6Z,9Z,12Z,15Z)/18:4(6Z,9Z,12Z,15Z))	0.89	1	0.85	-0.03
Dehydrotomatine	0.73	1	0.30	-0.28
Tin chloride (SnCl4)	0.64	1	0.82	0.02
PC(14:1(9Z)/22:2(13Z,16Z))	0.28	1	0.31	-0.16
PC(o-18:1(11Z)/18:2(9Z,12Z))	0.62	1	0.70	-0.09
Arginine	0	0.60	0.70	-0.07
Adenosine	0.11	1	0.40	-0.30
6-Deoxodolichosterone	0.81	1	0.33	-0.34
Niclosamide	0.41	1	0.07	-0.52
LysoPC(20:4(8Z,11Z,14Z,17Z))	0.02	1	0.11	-0.55
1,2,4,8-Tetraacetoxy-7-hydroxy-3-(4-hydroxyphenyl) dibenzofuran	0.66	1	0.34	-0.57
3-Deoxy-D-glycero-D-galacto-2-nonulosonic acid	0.04	1	0.37	-0.43
Glycerophosphocholine (38:6)	0.71	1	0.54	-0.13
N-(2-formyl-3-chlorophenyl) anthranilic acid	0.39	1	0.66	-0.04
5-Aminopentanoic acid	0.75	1	0.90	0.01
Narciclasine	0	0	0.38	0.19
Uric acid	0.93	1	0.34	-0.14
PE-NMe2(18:1(9Z)/18:1(9Z))	0.54	1	0.38	-0.15
PE(22:1(13Z)/15:0)	0.45	1	0.70	-0.07
[FA hydroxy(4:0)] 1,9S,11R,15S-tetrahydroxy-13E-prostaene	0.94	1	0.52	0.34
2,3-Butanediol glucoside	0.63	1	0.57	-0.24
Cer(d18:0/20:0)	0.84	1	0.27	0.26
LysoPE(18:0/0:0)	0.20	1	0.15	-0.44
Camelliagenin A	0.82	1	0.87	-0.11
LysoPC(16:0)	0	0.38	0.13	-0.37

FA amino(7:0) aminoheptanoic acid	0	0.22	0.24	-0.22
LysoPC(20:5(5Z,8Z,11Z,14Z,17Z))	0	1	0.99	0
PC(18:1(9Z)/20:1(11Z))	0.36	1	0.77	-0.05
Propanoylagmatine	0.96	1	0.76	-0.04
Lysyl-Tryptophan	0.85	1	0.67	-0.13
hydroxymethyl-dCDP	0.04	1	0.62	-0.09
Lactic acid	0.12	1	0.85	-0.07
N-[2-(4-Chloro-phenyl)-acetyl]-N'-(4,7-dimethyl-quinazolin-2-yl)- guanidine	0.08	1	0.90	-0.03
Deoxycytidine	0.26	1	0.78	-0.37
Glycerophosphocholine	0.05	1	0.09	-0.44
1-Octen-3-yl glucoside	0.61	1	0.90	-0.05
Di-2-propenyl hexasulfide	0.86	1	0.47	0.06
PC(20:3(8Z,11Z,14Z)/20:1(11Z))	0.88	1	0.74	0.10
PE(20:2(11Z,14Z)/15:0)	0.36	1	0.38	-0.29
N3-fumaramoyl-L-2,3-diaminopropanoate	0.13	1	0.74	-0.27
Leucyl-Proline	0.83	1	0.44	-0.25
(2Z,4'Z)-2-(5-Methylthio-4-penten-2-ynylidene)-1,6-dioxaspiro[4.4]non-3-ene	0	0.94	0.35	-0.15
Lenticin	1	1	0.77	-0.13
Tetrahexosylceramide (d18:1/16:0)	0.54	1	0.47	-0.17
N-(2-hydroxyhexacosanoyl)-eicosasphinganine-1-O-[D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.58	1	0.29	-0.16
PC(18:2(9Z,12Z)/18:2(9Z,12Z))	0.78	1	0.53	-0.10
Acetic acid	0.15	1	0.39	-0.13
[ST (20:4)] cholest-5-en-3beta-yl (15S-hydroperoxy-5Z,8Z,12E,14Z-eicosatetraenoate)	0.75	1	0.96	-0.01
Trihydroxy-3,6,7,4'-tetramethoxyflavone	0.01	1	0.45	-0.12
PC(20:3(8Z,11Z,14Z)/20:2(11Z,14Z))	0.70	1	0.33	-0.15
PG(16:0/16:0)	0.93	1	0.21	0.63
Dimethylarginine	0.18	1	0.40	-0.15
Hordatine B glucoside	0.50	1	0.16	0.29
N3-(4-methoxyfumaroyl)-L-2,3-diaminopropanoate	0.12	1	0.43	-0.27
PS(35:0)	0.47	1	0.39	-0.30
Glyceric acid	0.07	1	0.55	-0.57
PE(18:3(6Z,9Z,12Z)/22:0)	0.27	1	0.52	-0.12
PI(16:1(9Z)/18:1(11Z))	0.40	1	0.60	0.14
Glycerophosphocholine (52:4)	0.88	1	0.52	0.06
Psidinin C	0.70	1	0.22	-0.76
Methyl 3-hydroxybutyrate	0.97	1	0.62	-0.13
Gnidicin	0.14	1	0.74	-0.30
Isocitric acid	0.78	1	0.33	-0.28
6-Phosphonoglucono-D-lactone	0.03	1	0.85	-0.09
2,3-Dimethylbenzofuran	1	1	0.90	0.10
Sorbitan stearate	0.86	1	0.38	-0.45
cis-(homo)2aconitate	0	0.05	0.28	-0.09
Probenecid	0.71	1	0.63	0.15

Deoxyuridine triphosphate	0.08	1	0.98	0.02
SM(d18:0/16:1(9Z))	0.70	1	0.40	-0.13
PC(22:2(13Z,16Z)/16:1(9Z))	0.59	1	0.73	0.08
PE(P-18:1(9Z)/14:1(9Z))	0.78	1	0.39	-0.08
5-(4-Acetoxy-1-butynyl)-2,2'-bithiophene	0.52	1	0.66	0.14
PE(P-18:1(9Z)/20:5(5Z,8Z,11Z,14Z,17Z))	0.34	1	0.86	-0.06
Pyruvic acid	0.05	1	0.38	-0.47
Creatinine	0.92	1	0.40	-0.12
(2alpha,3alpha,5alpha,22R,23R)-2,3,22,23-Tetrahydroxy-25-methylergost-24(28)en-6-one	0.06	1	0.25	-0.43
PE-NMe(16:0/16:0)	0.81	1	0.88	-0.05
Pipecolic acid	0.98	1	0.87	0.09
PG(42:6)	0.56	1	0.26	-0.21
N-(2-hydroxytetracosanoyl)-4,8-sphingadienine	0.96	1	0.64	-0.17
Diisopropyl sulfide	0.56	1	0.63	0.16
Lumequoylacetone	0.82	1	0.78	-0.16
N-(2-hydroxyhexacosanoyl)-4R-hydroxyeicosasphinganine-1-O-[D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.71	1	0.66	-0.04
PE(18:3(6Z,9Z,12Z)/20:0)	0.30	1	0.59	-0.19
Prenol 30,32-dihydroxy-2b-methyl-bishomohopane	0.71	1	0.29	-0.77
2-Hydroxyundecanoate	0.99	1	0.94	-0.07
PC(18:3(6Z,9Z,12Z)/20:3(8Z,11Z,14Z))	0.92	1	0.53	-0.12
4'-Methoxy-2',3,7-trihydroxyisoflavanone	0.37	1	0.39	-0.12
Pyrrolidine	0.14	1	0.60	-0.07
Cer(d18:1/24:1(15Z))	0.90	1	0.37	-0.51
1,6-Naphthalenedisulfonic acid	0.09	1	0.39	-0.24
TG(41:0)	0.86	1	0.53	-0.11
CDP-Glycerophospholipid (38:5)	0.12	1	0.21	-0.30
Ganglioside GM3 (d18:1/24:1(15Z))	0.13	1	0.33	-0.24
SM(d18:0/16:0)	0.99	1	0.88	-0.07
Luteolin 7,4'-diglucuronide-3'-glucoside	0.37	1	0.16	-0.89
5,8-Dihydro-6-(4-methyl-3-pentenyl)-1,2,3,4-tetrathiocin	0	0	0.09	-0.15
4-Hydroxybutyric acid	0.40	1	0.77	0.23
PS(P-39:0)	0.25	1	0.25	-0.24
7-methylthioheptanaldoxime	0.38	1	0.24	-1.11
PC(18:2(9Z,12Z)/20:2(11Z,14Z))	0.86	1	0.39	-0.15
PI(18:2(9Z,12Z)/20:1(11Z))	0.57	1	0.42	-0.32
PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.89	1	0.67	0.19
UDP-N-acetylmuramoyl-L-Ala-D-gamma-Glu-6-carboxy-L-Lys-D-Ala	0.92	1	0.37	0.06
PI (36:0)	0.74	1	0.65	-0.15
Stenopalustroside A	0.48	1	0.70	-0.22
PC(20:3(8Z,11Z,14Z)/18:2(9Z,12Z))	0.81	1	0.35	-0.14
6-Hydroxyluteolin 6-xyloside	0.44	1	0.87	-0.07
Triradylglycerol (61:8)	0.90	1	0.70	0.04

TG(15:0/20:0/20:0)	0.70	1	0.89	-0.02
Piperidine	0.02	1	0.54	-0.07
Phosphocreatine	0.72	1	0.22	-0.50
5-Hexyltetrahydro-2-furanoctanoic acid	0.94	1	0.56	-0.21
Proline betaine	1	1	0.98	0.01
Iodixanol	0.51	1	0.07	-0.26
Triradylglycerol (65:6)	0.53	1	0.96	0
3'-O-Methyladenosine	0.36	1	0.23	-0.18
Kanzonol R	0.84	1	0.59	-0.30
5-(3-Buten-1-ynyl)-2,2'-bithienyl	0.87	1	0.44	0.10
1-Pyrrolinium	0.99	1	0.81	-0.11
SM(d18:1/18:1(11Z))	0.38	1	0.72	0.10
SM(d33:1)	0.72	1	0.63	-0.12
beta-D-Glucosyl-1,4-N-acetyl-D-glucosaminyldiphosphoundecaprenol	0.95	1	0.78	-0.02
2,2,4-Trimethyl-3-(4-fluorophenyl)-2H-1-benzopyran-7-ol acetate	0.84	1	0.57	0.13
Imidazoleacetic acid riboside	0.02	1	0.73	0.04

Table A2 Metabolome changes in the control group at time point 1 hour (after the LPS administration) compared to the baseline of this study, - 50 minutes. 1-way ANOVA results in the global.p values, and t-tests are the local / pairwise comparisons (local p)**.**

Metabolite	global.p	global.p.adj	Control 60/Control -50: local.p	Control 60/Control -50: ratio
Propionylcarnitine	0.12	1.00	0.02	-0.63
PC(14:1(9Z)/22:4(7Z,10Z,13Z,16Z))	0.04	1.00	0.03	-0.67
Kaempferol 3-rhamnosyl-(1->2)-[glucosyl-(1->3)-(4'''-p-coumaroylrhamnosyl)-(1->6)-galactoside]	0.03	1.00	0.03	0.34
Phosphocreatinine	0.45	1.00	0.06	0.43
L-Histidine	0.65	1.00	0.07	0.50
Hymenocardine	0.34	1.00	0.08	0.47
Chinomethionat	0.49	1.00	0.08	-0.94
Glycerophosphoinositol (41:6)	0.17	1.00	0.09	-0.84
N(6)-Trimethyl-L-lysine	0.01	1.00	0.09	-0.34
5'-Dehydroadenosine	0.02	1.00	0.09	-0.18
PE(P-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.36	1.00	0.10	-0.65
Moracin	0.12	1.00	0.10	-0.60
N-desmethylclarithromycin	0.38	1.00	0.11	0.25
Glutamine	0.01	1.00	0.12	0.25
CDP-Glycerophospholipid(40:6)	0.11	1.00	0.12	-0.30
PC(P-18:0/15:0)	0.21	1.00	0.12	1.10
[FA amino(8:0)] 3-amino-octanoic acid	0.00	0.38	0.13	-0.29
PI(O-20:0/14:1(9Z))	0.30	1.00	0.13	-0.41

Urea	0.11	1.00	0.13	-0.20
Triradylglycero (60:5)	0.23	1.00	0.14	-0.54
Prolyl-Gamma-glutamate	0.07	1.00	0.14	-0.19
[PR trimethyl,methyl,methyl(5:0/15:0)] 2,6,14-trimethyl-6,7-epoxy-10-methylene-9-(3-methylpent-4-enyl)-pentadec-2-ene	0.09	1.00	0.14	0.39
1-18:2-2-18:3-monogalactosyldiacylglycerol	0.41	1.00	0.15	-0.67
N(alpha)-t-Butoxycarbonyl-L-leucine	0.24	1.00	0.16	-0.99
Bruceoside A	0.01	1.00	0.17	-0.66
Phosphatidyl glycerophosphate (O-42:2)	0.49	1.00	0.17	-0.40
PS(41:2)	0.11	1.00	0.17	-0.32
PC(O-16:0/18:2(9Z,12Z))	0.10	1.00	0.18	-0.23
N-(2-hydroxy-tetracosanoyl)-1-beta-glucosyl-4E,6E-hexadecasphingadienine	0.54	1.00	0.18	-0.19
LysoPC(18:2(9Z,12Z))	0.00	0.01	0.19	-0.34
L-Carnitine	0.01	1.00	0.19	-0.12
PC(o-18:0/20:4(8Z,11Z,14Z,17Z))	0.35	1.00	0.21	-0.19
Epimedin K	0.52	1.00	0.22	0.13
Galactaric acid	0.32	1.00	0.22	-1.28
Iodipamide	0.73	1.00	0.23	-0.70
PE(P-18:1(9Z)/18:3(6Z,9Z,12Z))	0.27	1.00	0.24	-0.36
Ranunculin	0.00	0.12	0.24	-0.89
Spirolide B	0.60	1.00	0.25	-0.17
CDP-Glycerophospholipid (18:1(9Z)/18:2(9Z,12Z))	0.32	1.00	0.25	-0.37
PS(39:4)	0.75	1.00	0.25	-0.28
3-Amino-1,2,4-triazole	0.52	1.00	0.25	0.49
Di-2-thienyl disulfide	0.19	1.00	0.25	-0.27
[6]-Gingerdiol 3,5-diacetate	0.74	1.00	0.26	0.53
PC(o-16:1(9Z)/18:2(9Z,12Z))	0.24	1.00	0.26	-0.20
Glycerophosphocholine (40:6)	0.15	1.00	0.27	-0.31
PI(O-36:2)	0.25	1.00	0.27	-0.39
CDP-Glycerophospholipid (38:4)	0.25	1.00	0.27	-0.31
PC(18:1(9Z)/18:1(9Z))	0.09	1.00	0.27	-0.22
PE(24:1(15Z)/15:0)	0.17	1.00	0.29	-0.21
PC(o-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.45	1.00	0.30	-0.17
Prostaglandin F1a	0.31	1.00	0.30	0.69
Succinic acid semialdehyde	0.01	1.00	0.31	-0.17
3-O-Caffeoyl-4-O-methylquinic acid	0.06	1.00	0.31	-0.20
PC(o-18:2(9Z,12Z)/18:2(9Z,12Z))	0.23	1.00	0.31	-0.19
PC(P-18:1(11Z)/22:5(7Z,10Z,13Z,16Z,19Z))	0.57	1.00	0.31	-0.18
Glucose	0.00	0.47	0.32	-0.11
Oxolucidine B	0.00	0.02	0.32	-0.25
Beta-1,4-D-Mannosylchitobiosyldiphosphodolichol	0.19	1.00	0.33	-0.17
PE-NMe(18:1(9Z)/18:1(9Z))	0.36	1.00	0.33	-0.21
Artonin K	0.00	0.00	0.34	-0.11
5-Ethoxy-4,5-dihydro-2(3H)furanone	0.00	0.65	0.34	-0.15

Didodecyl thiobispropanoate	0.38	1.00	0.35	-0.55
Erucicoyl-EA	0.75	1.00	0.35	0.58
PC(16:0/16:0)	0.57	1.00	0.35	-0.26
Xanthine-8-carboxylate	0.10	1.00	0.36	-0.25
N-(tetracosanoyl)-sphinganine-1-O-[myo-inositol-1-phosphoryl-6-D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.56	1.00	0.37	0.14
PC(o-16:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.48	1.00	0.37	-0.17
hexane-6-keto-1,3,4,6-tetracarboxylate	0.25	1.00	0.38	0.77
Ganglioside GM3 (d18:0/20:0)	0.50	1.00	0.38	-0.10
Epoxymurin B	0.77	1.00	0.39	-0.49
carboxynorspermidine	0.59	1.00	0.39	0.27
PE(15:0/22:0)	0.33	1.00	0.40	-0.16
Glycerophosphoserine (55:0)	0.39	1.00	0.40	-0.17
PS(37:2)	0.34	1.00	0.40	-0.28
DG(43:5)	0.63	1.00	0.41	-0.18
3S-hydroxyhexacosanoyl-CoA	0.56	1.00	0.41	-0.16
Trihexosylceramide (d18:1/24:0)	0.84	1.00	0.42	0.12
PS(21:0/0:0)	0.00	0.27	0.42	-0.27
PE(P-18:1(9Z)/14:0)	0.32	1.00	0.44	0.11
Creatine	0.72	1.00	0.44	-0.29
L-Acetylcarnitine	0.01	1.00	0.44	-0.19
PC(o-16:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.63	1.00	0.45	-0.14
PE(18:4(6Z,9Z,12Z,15Z)/20:2(11Z,14Z))	0.75	1.00	0.45	0.44
Tartronate semialdehyde	0.37	1.00	0.45	0.40
SM(d34:2)	0.65	1.00	0.45	-0.16
Sedoheptulose	0.69	1.00	0.46	-0.16
PE(22:5(7Z,10Z,13Z,16Z,19Z)/16:0)	0.88	1.00	0.46	-0.43
LysoPE(20:0/0:0)	0.04	1.00	0.47	0.52
Diphenylcarbazide	0.65	1.00	0.47	0.41
PC(P-18:1(11Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.61	1.00	0.47	-0.18
CDP-DG(16:0/20:4(5Z,8Z,11Z,14Z))	0.14	1.00	0.47	-0.11
PE-NMe(16:0/18:1(9Z))	0.54	1.00	0.48	-0.39
PC(14:1(9Z)/20:2(11Z,14Z))	0.17	1.00	0.48	-0.15
L-Cyclo(alanylglycyl)	0.77	1.00	0.49	-0.48
PG(O-36:1)	0.26	1.00	0.49	-0.14
CDP-Glycerophospholipid (38:6)	0.67	1.00	0.49	-0.18
D-Sedoheptulose 7-phosphate	0.87	1.00	0.49	-0.17
PS(41:4)	0.44	1.00	0.50	-0.16
PE(18:4(6Z,9Z,12Z,15Z)/18:4(6Z,9Z,12Z,15Z))	0.89	1.00	0.51	0.12
Dehydrotomatine	0.73	1.00	0.51	-0.19
Tin chloride (SnCl4)	0.64	1.00	0.52	-0.05
PC(14:1(9Z)/22:2(13Z,16Z))	0.28	1.00	0.52	-0.12
PC(o-18:1(11Z)/18:2(9Z,12Z))	0.62	1.00	0.53	-0.14
Arginine	0.00	0.60	0.53	0.12
Adenosine	0.11	1.00	0.54	0.20

6-Deoxodolichosterone	0.81	1.00	0.54	-0.22
Niclosamide	0.41	1.00	0.55	-0.17
LysoPC(20:4(8Z,11Z,14Z,17Z))	0.02	1.00	0.55	-0.20
1,2,4,8-Tetraacetoxy-7-hydroxy-3-(4-hydroxyphenyl)dibenzofuran	0.66	1.00	0.56	-0.36
3-Deoxy-D-glycero-D-galacto-2-nonulosonic acid	0.04	1.00	0.56	-0.29
Glycerophosphocholine (38:6)	0.71	1.00	0.56	-0.13
N-(2-formyl-3-chlorophenyl)anthranilic acid	0.39	1.00	0.57	0.05
5-Aminopentanoic acid	0.75	1.00	0.57	0.06
Narciclasine	0.00	0.00	0.57	0.14
Uric acid	0.93	1.00	0.58	-0.10
PE-NMe2(18:1(9Z)/18:1(9Z))	0.54	1.00	0.59	-0.11
PE(22:1(13Z)/15:0)	0.45	1.00	0.59	-0.12
[FA hydroxy(4:0)] 1,9S,11R,15S-tetrahydroxy-13E-prostaene	0.94	1.00	0.59	0.31
2,3-Butanediol glucoside	0.63	1.00	0.60	-0.20
Cer(d18:0/20:0)	0.84	1.00	0.60	0.18
LysoPE(18:0/0:0)	0.20	1.00	0.60	-0.17
Camelliagenin A	0.82	1.00	0.60	-0.28
LysoPC(16:0)	0.00	0.38	0.60	-0.14
FA amino(7:0) aminoheptanoic acid	0.00	0.22	0.61	-0.09
LysoPC(20:5(5Z,8Z,11Z,14Z,17Z))	0.00	1.00	0.62	-0.14
PC(18:1(9Z)/20:1(11Z))	0.36	1.00	0.62	-0.13
Propanoylagmatine	0.96	1.00	0.63	-0.06
Lysyl-Tryptophan	0.85	1.00	0.64	-0.17
hydroxymethyl-dCDP	0.04	1.00	0.65	0.08
Lactic acid	0.12	1.00	0.65	0.28
N-[2-(4-Chloro-phenyl)-acetyl]-N'-(4,7-dimethyl-quinazolin-2-yl)- guanidine	0.08	1.00	0.65	-0.09
Deoxycytidine	0.26	1.00	0.65	0.77
Glycerophosphocholine	0.05	1.00	0.66	-0.13
1-Octen-3-yl glucoside	0.61	1.00	0.66	-0.23
Di-2-propenyl hexasulfide	0.86	1.00	0.66	0.07
PC(20:3(8Z,11Z,14Z)/20:1(11Z))	0.88	1.00	0.66	0.19
PE(20:2(11Z,14Z)/15:0)	0.36	1.00	0.67	-0.13
N3-fumaramoyl-L-2,3-diaminopropanoate	0.13	1.00	0.67	0.45
Leucyl-Proline	0.83	1.00	0.68	-0.13
(2Z,4'Z)-2-(5-Methylthio-4-penten-2-ynylidene)-1,6-dioxaspiro[4.4]non-3-ene	0.00	0.94	0.68	-0.06
Lenticin	1.00	1.00	0.68	-0.20
Tetrahexosylceramide (d18:1/16:0)	0.54	1.00	0.69	0.12
N-(2-hydroxyhexacosanoyl)-eicosasphinganine-1-O-[D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.58	1.00	0.69	-0.05
PC(18:2(9Z,12Z)/18:2(9Z,12Z))	0.78	1.00	0.69	-0.09
Acetic acid	0.15	1.00	0.69	-0.06
[ST (20:4)] cholest-5-en-3beta-yl (15S-hydroperoxy-5Z,8Z,12E,14Z-eicosatetraenoate)	0.75	1.00	0.70	0.07

Trihydroxy-3,6,7,4'-tetramethoxyflavone	0.01	1.00	0.70	0.08
PC(20:3(8Z,11Z,14Z)/20:2(11Z,14Z))	0.70	1.00	0.70	0.09
PG(16:0/16:0)	0.93	1.00	0.71	0.23
Dimethylarginine	0.18	1.00	0.71	0.07
Hordatine B glucoside	0.50	1.00	0.71	0.09
N3-(4-methoxyfumaroyl)-L-2,3-diaminopropanoate	0.12	1.00	0.72	0.19
PS(35:0)	0.47	1.00	0.72	-0.11
Glyceric acid	0.07	1.00	0.73	-0.26
PE(18:3(6Z,9Z,12Z)/22:0)	0.27	1.00	0.74	0.08
PI(16:1(9Z)/18:1(11Z))	0.40	1.00	0.74	0.09
Glycerophosphocholine (52:4)	0.88	1.00	0.75	0.04
Psidinin C	0.70	1.00	0.75	-0.17
Methyl 3-hydroxybutyrate	0.97	1.00	0.75	0.12
Gnidicin	0.14	1.00	0.75	0.30
Isocitric acid	0.78	1.00	0.75	0.09
6-Phosphonoglucono-D-lactone	0.03	1.00	0.76	0.20
2,3-Dimethylbenzofuran	1.00	1.00	0.76	0.24
Sorbitan stearate	0.86	1.00	0.76	-0.17
cis-(homo)2aconitate	0.00	0.05	0.76	0.03
Probenecid	0.71	1.00	0.76	-0.10
Deoxyuridine triphosphate	0.08	1.00	0.76	0.29
SM(d18:0/16:1(9Z))	0.70	1.00	0.77	-0.06
PC(22:2(13Z,16Z)/16:1(9Z))	0.59	1.00	0.78	-0.07
PE(P-18:1(9Z)/14:1(9Z))	0.78	1.00	0.78	0.03
5-(4-Acetoxy-1-butynyl)-2,2'-bithiophene	0.52	1.00	0.78	0.09
PE(P-18:1(9Z)/20:5(5Z,8Z,11Z,14Z,17Z))	0.34	1.00	0.79	-0.09
Pyruvic acid	0.05	1.00	0.80	-0.15
Creatinine	0.92	1.00	0.80	-0.04
(2alpha,3alpha,5alpha,22R,23R)-2,3,22,23-Tetrahydroxy-25-methylergost-24(28)en-6-one	0.06	1.00	0.81	-0.11
PE-NMe(16:0/16:0)	0.81	1.00	0.82	0.08
Pipecolic acid	0.98	1.00	0.82	0.13
PG(42:6)	0.56	1.00	0.83	-0.05
N-(2-hydroxytetracosanoyl)-4,8-sphingadienine	0.96	1.00	0.84	0.08
Diisopropyl sulfide	0.56	1.00	0.84	0.07
Lumequoylacetone	0.82	1.00	0.84	-0.13
N-(2-hydroxyhexacosanoyl)-4R-hydroxyeicosasphinganine-1-O-[D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.71	1.00	0.84	0.02
PE(18:3(6Z,9Z,12Z)/20:0)	0.30	1.00	0.86	0.07
Prenol 30,32-dihydroxy-2b-methyl-bishomohopane	0.71	1.00	0.86	-0.10
2-Hydroxyundecanoate	0.99	1.00	0.86	0.16
PC(18:3(6Z,9Z,12Z)/20:3(8Z,11Z,14Z))	0.92	1.00	0.86	-0.05
4'-Methoxy-2',3,7-trihydroxyisoflavanone	0.37	1.00	0.86	-0.02
Pyrrolidine	0.14	1.00	0.86	-0.02

Cer(d18:1/24:1(15Z))	0.90	1.00	0.86	-0.11
1,6-Naphthalenedisulfonic acid	0.09	1.00	0.87	-0.06
TG(41:0)	0.86	1.00	0.87	-0.04
CDP-Glycerophospholipid (38:5)	0.12	1.00	0.87	-0.04
Ganglioside GM3 (d18:1/24:1(15Z))	0.13	1.00	0.88	0.04
SM(d18:0/16:0)	0.99	1.00	0.88	0.07
Luteolin 7,4'-diglucuronide-3'-glucoside	0.37	1.00	0.88	-0.07
5,8-Dihydro-6-(4-methyl-3-pentenyl)-1,2,3,4-tetrathiocin	0.00	0.00	0.89	0.01
4-Hydroxybutyric acid	0.40	1.00	0.89	-0.11
PS(P-39:0)	0.25	1.00	0.89	-0.04
7-methylthioheptanaldoxime	0.38	1.00	0.89	-0.11
PC(18:2(9Z,12Z)/20:2(11Z,14Z))	0.86	1.00	0.90	0.03
PI(18:2(9Z,12Z)/20:1(11Z))	0.57	1.00	0.91	0.05
PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.89	1.00	0.91	-0.06
UDP-N-acetylmuramoyl-L-Ala-D-gamma-Glu-6-carboxy-L-Lys-D-Ala	0.92	1.00	0.91	0.01
PI (36:0)	0.74	1.00	0.91	-0.04
Stenopalustroside A	0.48	1.00	0.91	-0.06
PC(20:3(8Z,11Z,14Z)/18:2(9Z,12Z))	0.81	1.00	0.91	-0.02
6-Hydroxyluteolin 6-xyloside	0.44	1.00	0.91	0.05
Triradylglycerol (61:8)	0.90	1.00	0.92	-0.01
TG(15:0/20:0/20:0)	0.70	1.00	0.92	-0.02
Piperidine	0.02	1.00	0.93	-0.01
Phosphocreatine	0.72	1.00	0.93	-0.04
5-Hexyltetrahydro-2-furanoctanoic acid	0.94	1.00	0.94	-0.02
Proline betaine	1.00	1.00	0.94	0.06
Iodixanol	0.51	1.00	0.94	0.01
Triradylglycerol (65:6)	0.53	1.00	0.95	-0.01
3'-O-Methyladenosine	0.36	1.00	0.95	0.01
Kanzonol R	0.84	1.00	0.95	0.04
5-(3-Buten-1-ynyl)-2,2'-bithienyl	0.87	1.00	0.96	0.01
1-Pyrrolinium	0.99	1.00	0.96	0.03
SM(d18:1/18:1(11Z))	0.38	1.00	0.96	-0.02
SM(d33:1)	0.72	1.00	0.96	-0.02
beta-D-Glucosyl-1,4-N-acetyl-D-glucosaminyldiphosphoundecaprenol	0.95	1.00	0.97	0.00
2,2,4-Trimethyl-3-(4-fluorophenyl)-2H-1-benzopyran-7-ol acetate	0.84	1.00	0.98	0.00
Imidazoleacetic acid riboside	0.02	1.00	0.99	0.00

Table A3 Metabolome changes in the control group at time point 2 hours (after the LPS administration) compared to the baseline of this study, - 50 minutes. 1-way ANOVA results in the global.p values, and t-tests are the local / pairwise comparisons (local p).

Metabolite	global.p	global.p.adj	Control 120/Control -50: local.p	Control 120/Control -50: ratio
Propionylcarnitine	0.12	1.00	0.02	-0.62
PC(14:1(9Z)/22:4(7Z,10Z,13Z,16Z))	0.04	1.00	0.02	-0.78
Kaempferol 3-rhamnosyl-(1->2)-[glucosyl-(1->3)-(4'''-p-coumaroylrhamnosyl)-(1->6)-galactoside]	0.03	1.00	0.83	0.02
Phosphocreatinine	0.45	1.00	0.21	0.30
L-Histidine	0.65	1.00	0.15	1.11
Hymenocardine	0.34	1.00	0.48	0.22
Chinomethionat	0.49	1.00	0.34	-0.44
Glycerophosphoinositol (41:6)	0.17	1.00	0.06	-0.94
N(6)-Trimethyl-L-lysine	0.01	1.00	0.10	-0.32
5'-Dehydroadenosine	0.02	1.00	0.06	-0.21
PE(P-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.36	1.00	0.10	-0.58
Moracin	0.12	1.00	0.03	-0.97
N-desmethylclarithromycin	0.38	1.00	0.08	0.21
Glutamine	0.01	1.00	0.94	-0.01
CDP-Glycerophospholipid(40:6)	0.11	1.00	0.01	-0.45
PC(P-18:0/15:0)	0.21	1.00	0.65	0.36
[FA amino(8:0)] 3-amino-octanoic acid	0.00	0.38	0.07	-0.35
PI(O-20:0/14:1(9Z))	0.30	1.00	0.03	-0.60
Urea	0.11	1.00	0.06	-0.28
Triradylglycero (60:5)	0.23	1.00	0.47	0.21
Prolyl-Gamma-glutamate	0.07	1.00	0.05	-0.27
[PR trimethyl,methyl,methyl(5:0/15:0)] 2,6,14-trimethyl-6,7-epoxy-10-methylene-9-(3-methylpent-4-enyl)-pentadec-2-ene	0.09	1.00	0.10	-0.91
1-18:2-2-18:3-monogalactosyldiacylglycerol	0.41	1.00	0.40	-0.27
N(alpha)-t-Butoxycarbonyl-L-leucine	0.24	1.00	0.16	-0.98
Bruceoside A	0.01	1.00	0.04	-1.04
Phosphatidyl glycerophosphate (O-42:2)	0.49	1.00	0.29	-0.28
PS(41:2)	0.11	1.00	0.02	-0.66
PC(O-16:0/18:2(9Z,12Z))	0.10	1.00	0.04	-0.41
N-(2-hydroxy-tetracosanoyl)-1-beta-glucosyl-4E,6E-hexadecasphingadienine	0.54	1.00	0.12	-0.26
LysoPC(18:2(9Z,12Z))	0.00	0.01	0.03	-0.58
L-Carnitine	0.01	1.00	0.06	-0.19
PC(o-18:0/20:4(8Z,11Z,14Z,17Z))	0.35	1.00	0.15	-0.23
Epimedin K	0.52	1.00	0.67	0.03
Galactaric acid	0.32	1.00	0.44	-0.70
Iodipamide	0.73	1.00	0.34	-0.55
PE(P-18:1(9Z)/18:3(6Z,9Z,12Z))	0.27	1.00	0.07	-0.61
Ranunculin	0.00	0.12	0.05	-2.08

Spirolide B	0.60	1.00	0.40	-0.10
CDP-Glycerophospholipid (18:1(9Z)/18:2(9Z,12Z))	0.32	1.00	0.05	-0.58
PS(39:4)	0.75	1.00	0.39	-0.22
3-Amino-1,2,4-triazole	0.52	1.00	0.76	0.13
Di-2-thienyl disulfide	0.19	1.00	0.03	-0.57
[6]-Gingerdiol 3,5-diacetate	0.74	1.00	0.69	-0.23
PC(o-16:1(9Z)/18:2(9Z,12Z))	0.24	1.00	0.10	-0.25
Glycerophosphocholine (40:6)	0.15	1.00	0.03	-0.64
PI(O-36:2)	0.25	1.00	0.05	-0.75
CDP-Glycerophospholipid (38:4)	0.25	1.00	0.09	-0.46
PC(18:1(9Z)/18:1(9Z))	0.09	1.00	0.07	-0.38
PE(24:1(15Z)/15:0)	0.17	1.00	0.09	-0.34
PC(o-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.45	1.00	0.23	-0.17
Prostaglandin F1a	0.31	1.00	0.09	1.02
Succinic acid semialdehyde	0.01	1.00	0.62	-0.06
3-O-Caffeoyl-4-O-methylquinic acid	0.06	1.00	0.45	-0.14
PC(o-18:2(9Z,12Z)/18:2(9Z,12Z))	0.23	1.00	0.14	-0.22
PC(P-18:1(11Z)/22:5(7Z,10Z,13Z,16Z,19Z))	0.57	1.00	0.62	-0.09
Glucose	0.00	0.47	0.32	-0.11
Oxolucidine B	0.00	0.02	0.03	-0.57
Beta-1,4-D-Mannosylchitobiosyldiphosphodolichol	0.19	1.00	0.63	0.06
PE-NMe(18:1(9Z)/18:1(9Z))	0.36	1.00	0.04	-0.39
Artonin K	0.00	0.00	0.20	-0.16
5-Ethoxy-4,5-dihydro-2(3H)furanone	0.00	0.65	0.57	-0.08
Didodecyl thiobispropanoate	0.38	1.00	0.47	-0.34
Erucicoyl-EA	0.75	1.00	0.22	0.75
PC(16:0/16:0)	0.57	1.00	0.19	-0.36
Xanthine-8-carboxylate	0.10	1.00	0.02	-0.57
N-(tetracosanoyl)-sphinganine-1-O-[myo-inositol-1-phosphoryl-6-D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.56	1.00	0.08	0.20
PC(o-16:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.48	1.00	0.23	-0.19
hexane-6-keto-1,3,4,6-tetracarboxylate	0.25	1.00	0.19	0.79
Ganglioside GM3 (d18:0/20:0)	0.50	1.00	0.76	0.03
Epoxymurin B	0.77	1.00	0.08	-0.92
carboxynorspermidine	0.59	1.00	0.87	-0.05
PE(15:0/22:0)	0.33	1.00	0.22	-0.22
Glycerophosphoserine (55:0)	0.39	1.00	0.04	-0.37
PS(37:2)	0.34	1.00	0.05	-0.64
DG(43:5)	0.63	1.00	0.08	-0.35
3S-hydroxyhexacosanoyl-CoA	0.56	1.00	0.11	-0.29
Trihexosylceramide (d18:1/24:0)	0.84	1.00	0.75	0.04
PS(21:0/0:0)	0.00	0.27	0.13	-0.57
PE(P-18:1(9Z)/14:0)	0.32	1.00	0.59	0.06
Creatine	0.72	1.00	0.41	-0.31
L-Acetylcarnitine	0.01	1.00	0.13	-0.40

PC(o-16:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.63	1.00	0.29	-0.17
PE(18:4(6Z,9Z,12Z,15Z)/20:2(11Z,14Z))	0.75	1.00	0.26	0.50
Tartronate semialdehyde	0.37	1.00	0.52	-0.33
SM(d34:2)	0.65	1.00	0.25	-0.23
Sedoheptulose	0.69	1.00	0.53	-0.13
PE(22:5(7Z,10Z,13Z,16Z,19Z)/16:0)	0.88	1.00	0.42	-0.47
LysoPE(20:0/0:0)	0.04	1.00	0.49	0.36
Diphenylcarbazide	0.65	1.00	0.86	-0.11
PC(P-18:1(11Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.61	1.00	0.11	-0.35
CDP-DG(16:0/20:4(5Z,8Z,11Z,14Z))	0.14	1.00	0.34	-0.15
PE-NMe(16:0/18:1(9Z))	0.54	1.00	0.78	-0.16
PC(14:1(9Z)/20:2(11Z,14Z))	0.17	1.00	0.33	-0.19
L-Cyclo(alanylglycyl)	0.77	1.00	0.41	-0.59
PG(O-36:1)	0.26	1.00	0.24	-0.23
CDP-Glycerophospholipid (38:6)	0.67	1.00	0.23	-0.27
D-Sedoheptulose 7-phosphate	0.87	1.00	0.31	-0.23
PS(41:4)	0.44	1.00	0.09	-0.38
PE(18:4(6Z,9Z,12Z,15Z)/18:4(6Z,9Z,12Z,15Z))	0.89	1.00	0.35	0.10
Dehydrotomatine	0.73	1.00	0.18	-0.36
Tin chloride (SnCl4)	0.64	1.00	0.63	0.04
PC(14:1(9Z)/22:2(13Z,16Z))	0.28	1.00	0.23	-0.20
PC(o-18:1(11Z)/18:2(9Z,12Z))	0.62	1.00	0.43	-0.17
Arginine	0.00	0.60	0.45	-0.15
Adenosine	0.11	1.00	0.85	-0.05
6-Deoxodolichosterone	0.81	1.00	0.16	-0.42
Niclosamide	0.41	1.00	0.29	-0.28
LysoPC(20:4(8Z,11Z,14Z,17Z))	0.02	1.00	0.11	-0.47
1,2,4,8-Tetraacetoxy-7-hydroxy-3-(4-hydroxyphenyl)dibenzofuran	0.66	1.00	0.14	-0.84
3-Deoxy-D-glycero-D-galacto-2-nonulosonic acid	0.04	1.00	0.46	-0.34
Glycerophosphocholine (38:6)	0.71	1.00	0.18	-0.31
N-(2-formyl-3-chlorophenyl)anthranilic acid	0.39	1.00	0.90	0.01
5-Aminopentanoic acid	0.75	1.00	0.46	0.09
Narciclasine	0.00	0.00	0.02	-0.56
Uric acid	0.93	1.00	0.38	-0.11
PE-NMe2(18:1(9Z)/18:1(9Z))	0.54	1.00	0.36	-0.18
PE(22:1(13Z)/15:0)	0.45	1.00	0.32	-0.20
[FA hydroxy(4:0)] 1,9S,11R,15S-tetrahydroxy-13E-prostaene	0.94	1.00	1.00	0.00
2,3-Butanediol glucoside	0.63	1.00	0.35	-0.33
Cer(d18:0/20:0)	0.84	1.00	0.71	-0.11
LysoPE(18:0/0:0)	0.20	1.00	0.21	-0.36
Camelliagenin A	0.82	1.00	0.31	-0.52
LysoPC(16:0)	0.00	0.38	0.11	-0.38
FA amino(7:0) aminoheptanoic acid	0.00	0.22	0.08	-0.34

LysoPC(20:5(5Z,8Z,11Z,14Z,17Z))	0.00	1.00	0.08	-0.61
PC(18:1(9Z)/20:1(11Z))	0.36	1.00	0.17	-0.35
Propanoylagmatine	0.96	1.00	0.84	0.03
Lysyl-Tryptophan	0.85	1.00	0.70	-0.13
hydroxymethyl-dCDP	0.04	1.00	0.84	-0.03
Lactic acid	0.12	1.00	0.39	0.30
N-[2-(4-Chloro-phenyl)-acetyl]-N'-(4,7-dimethyl-quinazolin-2-yl)- guanidine	0.08	1.00	0.97	-0.01
Deoxycytidine	0.26	1.00	0.68	0.48
Glycerophosphocholine	0.05	1.00	0.17	-0.36
1-Octen-3-yl glucoside	0.61	1.00	0.26	-0.55
Di-2-propenyl hexasulfide	0.86	1.00	0.07	0.16
PC(20:3(8Z,11Z,14Z)/20:1(11Z))	0.88	1.00	0.74	-0.12
PE(20:2(11Z,14Z)/15:0)	0.36	1.00	0.48	-0.21
N3-fumaramoyl-L-2,3-diaminopropanoate	0.13	1.00	0.43	0.49
Leucyl-Proline	0.83	1.00	0.27	-0.36
(2Z,4'Z)-2-(5-Methylthio-4-penten-2-ynylidene)-1,6-dioxaspiro[4.4]non-3-ene	0.00	0.94	0.59	-0.08
Lenticin	1.00	1.00	0.64	-0.21
Tetrahexosylceramide (d18:1/16:0)	0.54	1.00	0.43	-0.17
N-(2-hydroxyhexacosanoyl)-eicosasphinganine-1-O-[D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.58	1.00	0.60	0.08
PC(18:2(9Z,12Z)/18:2(9Z,12Z))	0.78	1.00	0.62	-0.10
Acetic acid	0.15	1.00	0.41	-0.13
[ST (20:4)] cholest-5-en-3beta-yl (15S-hydroperoxy-5Z,8Z,12E,14Z-eicosatetraenoate)	0.75	1.00	0.81	-0.04
Trihydroxy-3,6,7,4'-tetramethoxyflavone	0.01	1.00	0.12	-0.25
PC(20:3(8Z,11Z,14Z)/20:2(11Z,14Z))	0.70	1.00	0.60	-0.10
PG(16:0/16:0)	0.93	1.00	0.63	0.27
Dimethylarginine	0.18	1.00	0.35	-0.18
Hordatine B glucoside	0.50	1.00	0.40	0.15
N3-(4-methoxyfumaroyl)-L-2,3-diaminopropanoate	0.12	1.00	0.44	0.24
PS(35:0)	0.47	1.00	0.07	-0.51
Glyceric acid	0.07	1.00	0.27	-1.02
PE(18:3(6Z,9Z,12Z)/22:0)	0.27	1.00	0.31	-0.26
PI(16:1(9Z)/18:1(11Z))	0.40	1.00	0.86	-0.05
Glycerophosphocholine (52:4)	0.88	1.00	0.52	0.07
Psidinin C	0.70	1.00	0.51	-0.34
Methyl 3-hydroxybutyrate	0.97	1.00	0.88	-0.04
Gnidicin	0.14	1.00	0.41	0.59
Isocitric acid	0.78	1.00	0.44	-0.23
6-Phosphonoglucono-D-lactone	0.03	1.00	0.22	0.52
2,3-Dimethylbenzofuran	1.00	1.00	1.00	0.00
Sorbitan stearate	0.86	1.00	0.22	-0.61
cis-(homo)2aconitate	0.00	0.05	0.72	-0.03
Probenecid	0.71	1.00	0.20	-0.36

Deoxyuridine triphosphate	0.08	1.00	0.59	0.39
SM(d18:0/16:1(9Z))	0.70	1.00	0.27	-0.17
PC(22:2(13Z,16Z)/16:1(9Z))	0.59	1.00	0.42	-0.23
PE(P-18:1(9Z)/14:1(9Z))	0.78	1.00	0.74	0.04
5-(4-Acetoxy-1-butynyl)-2,2'-bithiophene	0.52	1.00	0.66	-0.11
PE(P-18:1(9Z)/20:5(5Z,8Z,11Z,14Z,17Z))	0.34	1.00	0.16	-0.46
Pyruvic acid	0.05	1.00	0.87	-0.07
Creatinine	0.92	1.00	0.53	-0.09
(2alpha,3alpha,5alpha,22R,23R)-2,3,22,23-Tetrahydroxy-25-methylergost-24(28)en-6-one	0.06	1.00	0.24	-0.40
PE-NMe(16:0/16:0)	0.81	1.00	1.00	0.00
Pipecolic acid	0.98	1.00	0.89	0.08
PG(42:6)	0.56	1.00	0.33	-0.19
N-(2-hydroxytetracosanoyl)-4,8-sphingadienine	0.96	1.00	0.77	-0.11
Diisopropyl sulfide	0.56	1.00	0.46	-0.28
Lumequoylacetone	0.82	1.00	0.49	-0.37
N-(2-hydroxyhexacosanoyl)-4R-hydroxyeicosasphinganine-1-O-[D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.71	1.00	0.51	0.05
PE(18:3(6Z,9Z,12Z)/20:0)	0.30	1.00	0.16	-0.47
Prenol 30,32-dihydroxy-2b-methyl-bishomohopane	0.71	1.00	0.15	-0.86
2-Hydroxyundecanoate	0.99	1.00	0.74	-0.32
PC(18:3(6Z,9Z,12Z)/20:3(8Z,11Z,14Z))	0.92	1.00	0.74	-0.07
4'-Methoxy-2',3,7-trihydroxyisoflavanone	0.37	1.00	0.13	-0.17
Pyrrolidine	0.14	1.00	0.31	-0.12
Cer(d18:1/24:1(15Z))	0.90	1.00	0.80	0.15
1,6-Naphthalenedisulfonic acid	0.09	1.00	0.42	0.20
TG(41:0)	0.86	1.00	0.38	-0.17
CDP-Glycerophospholipid (38:5)	0.12	1.00	0.03	-0.59
Ganglioside GM3 (d18:1/24:1(15Z))	0.13	1.00	0.31	-0.22
SM(d18:0/16:0)	0.99	1.00	0.91	-0.06
Luteolin 7,4'-diglucuronide-3'-glucoside	0.37	1.00	0.18	-0.88
5,8-Dihydro-6-(4-methyl-3-pentenyl)-1,2,3,4-tetrathiocin	0.00	0.00	0.92	-0.01
4-Hydroxybutyric acid	0.40	1.00	0.85	-0.16
PS(P-39:0)	0.25	1.00	0.20	-0.29
7-methylthioheptanaldoxime	0.38	1.00	0.88	-0.13
PC(18:2(9Z,12Z)/20:2(11Z,14Z))	0.86	1.00	0.66	-0.09
PI(18:2(9Z,12Z)/20:1(11Z))	0.57	1.00	0.48	-0.27
PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.89	1.00	0.44	0.35
UDP-N-acetylmuramoyl-L-Ala-D-gamma-Glu-6-carboxy-L-Lys-D-Ala	0.92	1.00	0.96	0.00
PI (36:0)	0.74	1.00	0.30	-0.38
Stenopalustroside A	0.48	1.00	0.16	-0.80
PC(20:3(8Z,11Z,14Z)/18:2(9Z,12Z))	0.81	1.00	0.70	-0.07
6-Hydroxyluteolin 6-xyloside	0.44	1.00	0.29	-0.50

Triradylglycerol (61:8)	0.90	1.00	0.61	-0.06
TG(15:0/20:0/20:0)	0.70	1.00	0.32	0.15
Piperidine	0.02	1.00	0.34	-0.11
Phosphocreatine	0.72	1.00	0.23	-0.51
5-Hexyltetrahydro-2-furanoctanoic acid	0.94	1.00	0.51	-0.19
Proline betaine	1.00	1.00	0.83	-0.16
Iodixanol	0.51	1.00	0.51	-0.10
Triradylglycerol (65:6)	0.53	1.00	0.52	-0.06
3'-O-Methyladenosine	0.36	1.00	0.13	-0.23
Kanzonol R	0.84	1.00	0.17	-0.95
5-(3-Buten-1-ynyl)-2,2'-bithienyl	0.87	1.00	0.37	0.11
1-Pyrrolinium	0.99	1.00	0.64	-0.23
SM(d18:1/18:1(11Z))	0.38	1.00	0.23	-0.38
SM(d33:1)	0.72	1.00	0.49	-0.22
beta-D-Glucosyl-1,4-N-acetyl-D-glucosaminyldiphosphoundecaprenol	0.95	1.00	0.59	0.04
2,2,4-Trimethyl-3-(4-fluorophenyl)-2H-1-benzopyran-7-ol acetate	0.84	1.00	0.62	-0.07
Imidazoleacetic acid riboside	0.02	1.00	0.84	-0.02

Table A 4 Metabolome changes in the control group at time point 4 hours (after the LPS administration) compared to the baseline of this study, - 50 minutes. 1-way ANOVA results in the global p values, and t-tests are the local / pairwise comparisons (local p).

Metabolite	global.p	global.p.adj	Control 240/Control -50: local.p	Control 240/Control -50: ratio
Propionylcarnitine	0.12	1.00	0.06	-0.54
PC(14:1(9Z)/22:4(7Z,10Z,13Z,16Z))	0.04	1.00	0.01	-0.70
Kaempferol 3-rhamnosyl-(1->2)-[glucosyl-(1->3)-(4'''-p-coumaroylrhamnosyl)-(1->6)-galactoside]	0.03	1.00	0.80	-0.02
Phosphocreatinine	0.45	1.00	0.63	0.13
L-Histidine	0.65	1.00	0.52	0.46
Hymenocardine	0.34	1.00	0.08	0.46
Chinomethionat	0.49	1.00	0.22	-0.55
Glycerophosphoinositol (41:6)	0.17	1.00	0.08	-0.93
N(6)-Trimethyl-L-lysine	0.01	1.00	0.10	-0.32
5'-Dehydroadenosine	0.02	1.00	0.13	0.29
PE(P-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.36	1.00	0.18	-0.42
Moracin	0.12	1.00	0.03	-0.82
N-desmethylclarithromycin	0.38	1.00	0.38	0.10
Glutamine	0.01	1.00	0.00	-0.47
CDP-Glycerophospholipid(40:6)	0.11	1.00	0.17	-0.27
PC(P-18:0/15:0)	0.21	1.00	0.09	1.13
[FA amino(8:0)] 3-amino-octanoic acid	0.00	0.38	0.01	-0.61

PI(O-20:0/14:1(9Z))	0.30	1.00	0.45	-0.23
Urea	0.11	1.00	0.06	-0.29
Triradylglycero (60:5)	0.23	1.00	0.68	0.11
Prolyl-Gamma-glutamate	0.07	1.00	0.39	0.17
[PR trimethyl,methyl,methyl(5:0/15:0)]	0.09	1.00	0.77	-0.10
2,6,14-trimethyl-6,7-epoxy-10-methylene-9-(3-methylpent-4-enyl)-pentadec-2-ene				
1-18:2-2-18:3-monogalactosyldiacylglycerol	0.41	1.00	0.07	-0.69
N(alpha)-t-Butoxycarbonyl-L-leucine	0.24	1.00	0.19	-0.91
Bruceoside A	0.01	1.00	0.00	-1.92
Phosphatidyl glycerophosphate (O-42:2)	0.49	1.00	0.04	-0.57
PS(41:2)	0.11	1.00	0.41	-0.26
PC(O-16:0/18:2(9Z,12Z))	0.10	1.00	0.03	-0.38
N-(2-hydroxy-tetracosanoyl)-1-beta-glucosyl-4E,6E-hexadecasphingadienine	0.54	1.00	0.15	-0.22
LysoPC(18:2(9Z,12Z))	0.00	0.01	0.00	-1.15
L-Carnitine	0.01	1.00	0.00	-0.42
PC(o-18:0/20:4(8Z,11Z,14Z,17Z))	0.35	1.00	0.03	-0.29
Epimedin K	0.52	1.00	0.21	0.10
Galactaric acid	0.32	1.00	0.13	-1.79
Iodipamide	0.73	1.00	0.81	-0.16
PE(P-18:1(9Z)/18:3(6Z,9Z,12Z))	0.27	1.00	0.77	-0.10
Ranunculin	0.00	0.12	0.47	-0.62
Spirolide B	0.60	1.00	0.86	-0.02
CDP-Glycerophospholipid (18:1(9Z)/18:2(9Z,12Z))	0.32	1.00	0.54	-0.19
PS(39:4)	0.75	1.00	0.51	-0.17
3-Amino-1,2,4-triazole	0.52	1.00	0.68	-0.17
Di-2-thienyl disulfide	0.19	1.00	0.11	-0.40
[6]-Gingerdiol 3,5-diacetate	0.74	1.00	0.94	-0.04
PC(o-16:1(9Z)/18:2(9Z,12Z))	0.24	1.00	0.06	-0.30
Glycerophosphocholine (40:6)	0.15	1.00	0.57	-0.20
PI(O-36:2)	0.25	1.00	0.36	-0.34
CDP-Glycerophospholipid (38:4)	0.25	1.00	0.48	-0.23
PC(18:1(9Z)/18:1(9Z))	0.09	1.00	0.03	-0.41
PE(24:1(15Z)/15:0)	0.17	1.00	0.03	-0.39
PC(o-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.45	1.00	0.07	-0.23
Prostaglandin F1a	0.31	1.00	0.22	0.83
Succinic acid semialdehyde	0.01	1.00	0.00	-0.62
3-O-Caffeoyl-4-O-methylquinic acid	0.06	1.00	0.08	0.46
PC(o-18:2(9Z,12Z)/18:2(9Z,12Z))	0.23	1.00	0.04	-0.30
PC(P-18:1(11Z)/22:5(7Z,10Z,13Z,16Z,19Z))	0.57	1.00	0.17	-0.20
Glucose	0.00	0.47	0.02	0.36
Oxolucidine B	0.00	0.02	0.00	-1.04
Beta-1,4-D-Mannosylchitobiosyldiphosphodolichol	0.19	1.00	0.30	0.12
PE-NMe(18:1(9Z)/18:1(9Z))	0.36	1.00	0.58	-0.15
Artonin K	0.00	0.00	0.01	0.53

5-Ethoxy-4,5-dihydro-2(3H)furanone	0.00	0.65	0.42	0.10
Didodecyl thiobispropanoate	0.38	1.00	0.13	-0.84
Erucicoyl-EA	0.75	1.00	0.98	-0.01
PC(16:0/16:0)	0.57	1.00	0.07	-0.42
Xanthine-8-carboxylate	0.10	1.00	0.17	-0.38
N-(tetracosanoyl)-sphinganine-1-O-[myo-inositol-1-phosphoryl-6-D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.56	1.00	0.91	0.01
PC(o-16:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.48	1.00	0.10	-0.25
hexane-6-keto-1,3,4,6-tetracarboxylate	0.25	1.00	0.02	1.41
Ganglioside GM3 (d18:0/20:0)	0.50	1.00	0.81	-0.02
Epoxymurin B	0.77	1.00	0.73	-0.22
carboxynorspermidine	0.59	1.00	0.65	-0.15
PE(15:0/22:0)	0.33	1.00	0.10	-0.25
Glycerophosphoserine (55:0)	0.39	1.00	0.35	-0.14
PS(37:2)	0.34	1.00	0.59	-0.23
DG(43:5)	0.63	1.00	0.78	-0.08
3S-hydroxyhexacosanoyl-CoA	0.56	1.00	0.43	-0.14
Trihexosylceramide (d18:1/24:0)	0.84	1.00	0.18	0.16
PS(21:0/0:0)	0.00	0.27	0.01	-1.43
PE(P-18:1(9Z)/14:0)	0.32	1.00	0.17	-0.20
Creatine	0.72	1.00	0.85	0.06
L-Acetylcarnitine	0.01	1.00	0.01	-0.73
PC(o-16:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.63	1.00	0.13	-0.22
PE(18:4(6Z,9Z,12Z,15Z)/20:2(11Z,14Z))	0.75	1.00	0.24	0.50
Tartronate semialdehyde	0.37	1.00	0.40	-0.45
SM(d34:2)	0.65	1.00	0.13	-0.25
Sedoheptulose	0.69	1.00	0.08	-0.30
PE(22:5(7Z,10Z,13Z,16Z,19Z)/16:0)	0.88	1.00	0.88	-0.09
LysoPE(20:0/0:0)	0.04	1.00	0.05	-1.51
Diphenylcarbazide	0.65	1.00	0.91	0.07
PC(P-18:1(11Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.61	1.00	0.93	-0.03
CDP-DG(16:0/20:4(5Z,8Z,11Z,14Z))	0.14	1.00	0.05	-0.30
PE-NMe(16:0/18:1(9Z))	0.54	1.00	0.28	-0.63
PC(14:1(9Z)/20:2(11Z,14Z))	0.17	1.00	0.11	-0.31
L-Cyclo(alanylglycyl)	0.77	1.00	0.37	-0.65
PG(O-36:1)	0.26	1.00	0.07	-0.30
CDP-Glycerophospholipid (38:6)	0.67	1.00	0.83	-0.05
D-Sedoheptulose 7-phosphate	0.87	1.00	0.47	-0.16
PS(41:4)	0.44	1.00	0.76	-0.08
PE(18:4(6Z,9Z,12Z,15Z)/18:4(6Z,9Z,12Z,15Z))	0.89	1.00	0.57	0.07
Dehydrotomatine	0.73	1.00	0.65	-0.12
Tin chloride (SnCl4)	0.64	1.00	0.48	-0.05
PC(14:1(9Z)/22:2(13Z,16Z))	0.28	1.00	0.03	-0.33
PC(o-18:1(11Z)/18:2(9Z,12Z))	0.62	1.00	0.25	-0.27
Arginine	0.00	0.60	0.02	-0.48

Adenosine	0.11	1.00	0.45	0.23
6-Deoxodolichosterone	0.81	1.00	0.42	-0.25
Niclosamide	0.41	1.00	0.14	-0.45
LysoPC(20:4(8Z,11Z,14Z,17Z))	0.02	1.00	0.00	-1.20
1,2,4,8-Tetraacetoxy-7-hydroxy-3-(4-hydroxyphenyl)dibenzofuran	0.66	1.00	0.99	0.01
3-Deoxy-D-glycero-D-galacto-2-nonulosonic acid	0.04	1.00	0.16	0.78
Glycerophosphocholine (38:6)	0.71	1.00	0.69	-0.10
N-(2-formyl-3-chlorophenyl)anthranilic acid	0.39	1.00	0.14	0.11
5-Aminopentanoic acid	0.75	1.00	0.70	0.05
Narciclasine	0.00	0.00	0.00	-0.90
Uric acid	0.93	1.00	0.74	-0.04
PE-NMe2(18:1(9Z)/18:1(9Z))	0.54	1.00	0.09	-0.27
PE(22:1(13Z)/15:0)	0.45	1.00	0.22	-0.21
[FA hydroxy(4:0)] 1,9S,11R,15S-tetrahydroxy-13E-prostaene	0.94	1.00	0.76	0.21
2,3-Butanediol glucoside	0.63	1.00	0.56	-0.20
Cer(d18:0/20:0)	0.84	1.00	0.91	-0.03
LysoPE(18:0/0:0)	0.20	1.00	0.05	-0.52
Camelliagenin A	0.82	1.00	0.26	-0.82
LysoPC(16:0)	0.00	0.38	0.00	-0.76
FA amino(7:0) aminoheptanoic acid	0.00	0.22	0.00	-0.71
LysoPC(20:5(5Z,8Z,11Z,14Z,17Z))	0.00	1.00	0.04	-0.71
PC(18:1(9Z)/20:1(11Z))	0.36	1.00	0.06	-0.43
Propanoylagmatine	0.96	1.00	0.87	0.02
Lysyl-Tryptophan	0.85	1.00	0.95	0.02
hydroxymethyl-dCDP	0.04	1.00	0.00	-0.53
Lactic acid	0.12	1.00	0.05	1.00
N-[2-(4-Chloro-phenyl)-acetyl]-N'-(4,7-dimethyl-quinazolin-2-yl)- guanidine	0.08	1.00	0.04	0.54
Deoxycytidine	0.26	1.00	0.12	2.04
Glycerophosphocholine	0.05	1.00	0.02	-0.68
1-Octen-3-yl glucoside	0.61	1.00	0.39	-0.47
Di-2-propenyl hexasulfide	0.86	1.00	0.21	0.10
PC(20:3(8Z,11Z,14Z)/20:1(11Z))	0.88	1.00	0.66	-0.14
PE(20:2(11Z,14Z)/15:0)	0.36	1.00	0.13	-0.46
N3-fumaramoyl-L-2,3-diaminopropanoate	0.13	1.00	0.07	1.46
Leucyl-Proline	0.83	1.00	0.35	-0.30
(2Z,4'Z)-2-(5-Methylthio-4-penten-2-ynylidene)-1,6-dioxaspiro[4.4]non-3-ene	0.00	0.94	0.05	0.45
Lenticin	1.00	1.00	0.71	-0.17
Tetrahexosylceramide (d18:1/16:0)	0.54	1.00	0.14	-0.31
N-(2-hydroxyhexacosanoyl)-eicosasphinganine-1-O-[D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.58	1.00	0.66	0.06
PC(18:2(9Z,12Z)/18:2(9Z,12Z))	0.78	1.00	0.22	-0.20
Acetic acid	0.15	1.00	0.13	0.30

[ST (20:4)] cholest-5-en-3beta-yl (15S-hydroperoxy-5Z,8Z,12E,14Z-eicosatetraenoate)	0.75	1.00	0.22	-0.17
Trihydroxy-3,6,7,4'-tetramethoxyflavone	0.01	1.00	0.04	0.39
PC(20:3(8Z,11Z,14Z)/20:2(11Z,14Z))	0.70	1.00	0.28	-0.17
PG(16:0/16:0)	0.93	1.00	0.83	0.14
Dimethylarginine	0.18	1.00	0.11	-0.28
Hordatine B glucoside	0.50	1.00	0.94	-0.01
N3-(4-methoxyfumaroyl)-L-2,3-diaminopropanoate	0.12	1.00	0.08	0.81
PS(35:0)	0.47	1.00	0.64	-0.15
Glyceric acid	0.07	1.00	0.02	1.29
PE(18:3(6Z,9Z,12Z)/22:0)	0.27	1.00	0.08	-0.45
PI(16:1(9Z)/18:1(11Z))	0.40	1.00	0.43	0.25
Glycerophosphocholine (52:4)	0.88	1.00	0.39	0.09
Psidinin C	0.70	1.00	0.52	-0.33
Methyl 3-hydroxybutyrate	0.97	1.00	0.81	0.07
Gnidicin	0.14	1.00	0.06	1.38
Isocitric acid	0.78	1.00	0.64	-0.14
6-Phosphonoglucono-D-lactone	0.03	1.00	0.02	1.09
2,3-Dimethylbenzofuran	1.00	1.00	0.95	-0.05
Sorbitan stearate	0.86	1.00	0.58	-0.34
cis-(homo)2aconitate	0.00	0.05	0.01	0.31
Probenecid	0.71	1.00	0.49	-0.17
Deoxyuridine triphosphate	0.08	1.00	0.06	1.62
SM(d18:0/16:1(9Z))	0.70	1.00	0.08	-0.23
PC(22:2(13Z,16Z)/16:1(9Z))	0.59	1.00	0.24	-0.30
PE(P-18:1(9Z)/14:1(9Z))	0.78	1.00	0.44	-0.09
5-(4-Acetoxy-1-butynyl)-2,2'-bithiophene	0.52	1.00	0.11	-0.49
PE(P-18:1(9Z)/20:5(5Z,8Z,11Z,14Z,17Z))	0.34	1.00	0.72	-0.13
Pyruvic acid	0.05	1.00	0.12	0.89
Creatinine	0.92	1.00	0.51	-0.09
(2alpha,3alpha,5alpha,22R,23R)-2,3,22,23-Tetrahydroxy-25-methylergost-24(28)en-6-one	0.06	1.00	0.01	-1.03
PE-NMe(16:0/16:0)	0.81	1.00	0.29	-0.37
Pipecolic acid	0.98	1.00	0.95	-0.03
PG(42:6)	0.56	1.00	0.91	0.03
N-(2-hydroxytetracosanoyl)-4,8-sphingadienine	0.96	1.00	0.69	0.14
Diisopropyl sulfide	0.56	1.00	0.27	-0.46
Lumequoylacetone	0.82	1.00	0.81	0.18
N-(2-hydroxyhexacosanoyl)-4R-hydroxyeicosasphinganine-1-O-[D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.71	1.00	0.59	-0.05
PE(18:3(6Z,9Z,12Z)/20:0)	0.30	1.00	0.43	-0.22
Prenol 30,32-dihydroxy-2b-methyl-bishomohopane	0.71	1.00	0.83	-0.13
2-Hydroxyundecanoate	0.99	1.00	0.77	-0.26

PC(18:3(6Z,9Z,12Z)/20:3(8Z,11Z,14Z))	0.92	1.00	0.59	-0.09
4'-Methoxy-2',3,7-trihydroxyisoflavanone	0.37	1.00	0.37	0.12
Pyrrolidine	0.14	1.00	0.70	-0.06
Cer(d18:1/24:1(15Z))	0.90	1.00	0.53	-0.33
1,6-Naphthalenedisulfonic acid	0.09	1.00	0.07	0.52
TG(41:0)	0.86	1.00	0.19	-0.23
CDP-Glycerophospholipid (38:5)	0.12	1.00	0.74	-0.09
Ganglioside GM3 (d18:1/24:1(15Z))	0.13	1.00	0.70	-0.08
SM(d18:0/16:0)	0.99	1.00	0.76	0.12
Luteolin 7,4'-diglucuronide-3'-glucoside	0.37	1.00	0.86	-0.11
5,8-Dihydro-6-(4-methyl-3-pentenyl)-1,2,3,4-tetrathiocin	0.00	0.00	0.00	0.31
4-Hydroxybutyric acid	0.40	1.00	0.20	1.12
PS(P-39:0)	0.25	1.00	0.14	-0.31
7-methylthioheptanaldoxime	0.38	1.00	0.34	-0.81
PC(18:2(9Z,12Z)/20:2(11Z,14Z))	0.86	1.00	0.27	-0.19
PI(18:2(9Z,12Z)/20:1(11Z))	0.57	1.00	0.85	-0.08
PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.89	1.00	0.63	0.21
UDP-N-acetylmuramoyl-L-Ala-D-gamma-Glu-6-carboxy-L-Lys-D-Ala	0.92	1.00	0.73	0.03
PI (36:0)	0.74	1.00	1.00	0.00
Stenopalustroside A	0.48	1.00	0.70	-0.23
PC(20:3(8Z,11Z,14Z)/18:2(9Z,12Z))	0.81	1.00	0.28	-0.16
6-Hydroxyluteolin 6-xyloside	0.44	1.00	0.47	-0.36
Triradylglycerol (61:8)	0.90	1.00	0.49	0.09
TG(15:0/20:0/20:0)	0.70	1.00	0.76	0.04
Piperidine	0.02	1.00	0.65	0.07
Phosphocreatine	0.72	1.00	0.49	-0.23
5-Hexyltetrahydro-2-furanoctanoic acid	0.94	1.00	0.78	-0.09
Proline betaine	1.00	1.00	0.74	-0.24
Iodixanol	0.51	1.00	0.60	-0.07
Triradylglycerol (65:6)	0.53	1.00	0.28	0.11
3'-O-Methyladenosine	0.36	1.00	0.70	-0.05
Kanzonol R	0.84	1.00	0.92	-0.07
5-(3-Buten-1-ynyl)-2,2'-bithienyl	0.87	1.00	0.98	0.00
1-Pyrrolinium	0.99	1.00	0.66	-0.21
SM(d18:1/18:1(11Z))	0.38	1.00	0.24	-0.35
SM(d33:1)	0.72	1.00	0.11	-0.41
beta-D-Glucosyl-1,4-N-acetyl-D-glucosaminyldiphosphoundecaprenol	0.95	1.00	0.71	0.03
2,2,4-Trimethyl-3-(4-fluorophenyl)-2H-1-benzopyran-7-ol acetate	0.84	1.00	0.22	-0.21
Imidazoleacetic acid riboside	0.02	1.00	0.01	0.40

Table A 5 Metabolome changes in the control group at time point 8 hours (after the LPS administration) compared to the baseline of this study, - 50 minutes. 1-way ANOVA results in the global.p values, and t-tests are the local / pairwise comparisons (local p).

Metabolite	global.p	global.p.adj	Control 480/Control -50: local.p	Control 480/Control -50: ratio
Propionylcarnitine	0.12	1.00	0.13	-0.52
PC(14:1(9Z)/22:4(7Z,10Z,13Z,16Z))	0.04	1.00	0.01	-0.84
Kaempferol 3-rhamnosyl-(1->2)-[glucosyl-(1->3)-(4'''-p-coumaroylrhamnosyl)-(1->6)-galactoside]	0.03	1.00	0.15	0.13
Phosphocreatinine	0.45	1.00	0.46	0.17
L-Histidine	0.65	1.00	0.22	1.05
Hymenocardine	0.34	1.00	0.87	0.05
Chinomethionat	0.49	1.00	0.21	-0.58
Glycerophosphoinositol (41:6)	0.17	1.00	0.03	-1.19
N(6)-Trimethyl-L-lysine	0.01	1.00	0.02	-0.54
5'-Dehydroadenosine	0.02	1.00	0.50	0.11
PE(P-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.36	1.00	0.04	-0.72
Moracin	0.12	1.00	0.03	-0.76
N-desmethylclarithromycin	0.38	1.00	0.53	0.07
Glutamine	0.01	1.00	0.73	-0.05
CDP-Glycerophospholipid(40:6)	0.11	1.00	0.00	-0.60
PC(P-18:0/15:0)	0.21	1.00	0.45	0.58
[FA amino(8:0)] 3-amino-octanoic acid	0.00	0.38	0.00	-0.63
PI(O-20:0/14:1(9Z))	0.30	1.00	0.04	-0.50
Urea	0.11	1.00	0.04	-0.31
Triradylglycero (60:5)	0.23	1.00	0.51	0.20
Prolyl-Gamma-glutamate	0.07	1.00	0.80	0.04
[PR trimethyl,methyl,methyl(5:0/15:0)] 2,6,14-trimethyl-6,7-epoxy-10-methylene-9-(3-methylpent-4-enyl)-pentadec-2-ene	0.09	1.00	0.83	0.09
1-18:2-2-18:3-monogalactosyldiacylglycerol	0.41	1.00	0.09	-0.53
N(alpha)-t-Butoxycarbonyl-L-leucine	0.24	1.00	0.17	-0.93
Bruceoside A	0.01	1.00	0.01	-1.63
Phosphatidyl glycerophosphate (O-42:2)	0.49	1.00	0.10	-0.45
PS(41:2)	0.11	1.00	0.01	-0.71
PC(O-16:0/18:2(9Z,12Z))	0.10	1.00	0.01	-0.50
N-(2-hydroxy-tetracosanoyl)-1-beta-glucosyl-4E,6E-hexadecasphingadienine	0.54	1.00	0.09	-0.25
LysoPC(18:2(9Z,12Z))	0.00	0.01	0.00	-1.05
L-Carnitine	0.01	1.00	0.02	-0.23
PC(o-18:0/20:4(8Z,11Z,14Z,17Z))	0.35	1.00	0.02	-0.34
Epimedin K	0.52	1.00	0.76	0.03
Galactaric acid	0.32	1.00	0.28	-1.06
Iodipamide	0.73	1.00	0.13	-0.88
PE(P-18:1(9Z)/18:3(6Z,9Z,12Z))	0.27	1.00	0.05	-0.62
Ranunculin	0.00	0.12	0.07	1.30

Spirolide B	0.60	1.00	0.46	-0.09
CDP-Glycerophospholipid (18:1(9Z)/18:2(9Z,12Z))	0.32	1.00	0.04	-0.59
PS(39:4)	0.75	1.00	0.03	-0.46
3-Amino-1,2,4-triazole	0.52	1.00	0.96	0.02
Di-2-thienyl disulfide	0.19	1.00	0.09	-0.38
[6]-Gingerdiol 3,5-diacetate	0.74	1.00	0.76	0.17
PC(o-16:1(9Z)/18:2(9Z,12Z))	0.24	1.00	0.01	-0.42
Glycerophosphocholine (40:6)	0.15	1.00	0.01	-0.79
PI(O-36:2)	0.25	1.00	0.02	-0.84
CDP-Glycerophospholipid (38:4)	0.25	1.00	0.01	-0.68
PC(18:1(9Z)/18:1(9Z))	0.09	1.00	0.01	-0.51
PE(24:1(15Z)/15:0)	0.17	1.00	0.02	-0.47
PC(o-18:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.45	1.00	0.03	-0.29
Prostaglandin F1a	0.31	1.00	0.54	0.39
Succinic acid semialdehyde	0.01	1.00	0.08	-0.28
3-O-Caffeoyl-4-O-methylquinic acid	0.06	1.00	0.32	0.26
PC(o-18:2(9Z,12Z)/18:2(9Z,12Z))	0.23	1.00	0.01	-0.40
PC(P-18:1(11Z)/22:5(7Z,10Z,13Z,16Z,19Z))	0.57	1.00	0.04	-0.38
Glucose	0.00	0.47	0.37	0.11
Oxolucidine B	0.00	0.02	0.00	-1.08
Beta-1,4-D-Mannosylchitobiosyldiphosphodolichol	0.19	1.00	0.23	0.14
PE-NMe(18:1(9Z)/18:1(9Z))	0.36	1.00	0.00	-0.51
Artonin K	0.00	0.00	0.13	0.24
5-Ethoxy-4,5-dihydro-2(3H)furanone	0.00	0.65	0.02	-0.44
Didodecyl thiobispropanoate	0.38	1.00	0.07	-1.13
Erucicoyl-EA	0.75	1.00	0.73	0.26
PC(16:0/16:0)	0.57	1.00	0.17	-0.33
Xanthine-8-carboxylate	0.10	1.00	0.81	-0.06
N-(tetracosanoyl)-sphinganine-1-O-[myo-inositol-1-phosphoryl-6-D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.56	1.00	0.18	0.12
PC(o-16:1(9Z)/20:4(8Z,11Z,14Z,17Z))	0.48	1.00	0.03	-0.35
hexane-6-keto-1,3,4,6-tetracarboxylate	0.25	1.00	0.09	0.95
Ganglioside GM3 (d18:0/20:0)	0.50	1.00	0.70	-0.05
Epoxymurin B	0.77	1.00	0.10	-0.74
carboxynorspermidine	0.59	1.00	0.38	0.28
PE(15:0/22:0)	0.33	1.00	0.03	-0.36
Glycerophosphoserine (55:0)	0.39	1.00	0.28	-0.15
PS(37:2)	0.34	1.00	0.01	-0.83
DG(43:5)	0.63	1.00	0.02	-0.35
3S-hydroxyhexacosanoyl-CoA	0.56	1.00	0.02	-0.33
Trihexosylceramide (d18:1/24:0)	0.84	1.00	0.87	0.02
PS(21:0/0:0)	0.00	0.27	0.01	-1.10
PE(P-18:1(9Z)/14:0)	0.32	1.00	0.61	0.05
Creatine	0.72	1.00	0.76	-0.10
L-Acetylcarnitine	0.01	1.00	0.00	-1.00

PC(o-16:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.63	1.00	0.07	-0.29
PE(18:4(6Z,9Z,12Z,15Z)/20:2(11Z,14Z))	0.75	1.00	0.89	-0.07
Tartronate semialdehyde	0.37	1.00	0.18	-0.81
SM(d34:2)	0.65	1.00	0.22	-0.21
Sedoheptulose	0.69	1.00	0.16	-0.28
PE(22:5(7Z,10Z,13Z,16Z,19Z)/16:0)	0.88	1.00	0.19	-0.62
LysoPE(20:0/0:0)	0.04	1.00	0.03	-1.98
Diphenylcarbazide	0.65	1.00	0.49	-0.47
PC(P-18:1(11Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	0.61	1.00	0.02	-0.38
CDP-DG(16:0/20:4(5Z,8Z,11Z,14Z))	0.14	1.00	0.04	-0.36
PE-NMe(16:0/18:1(9Z))	0.54	1.00	0.18	-0.85
PC(14:1(9Z)/20:2(11Z,14Z))	0.17	1.00	0.02	-0.52
L-Cyclo(alanylglycyl)	0.77	1.00	0.38	-0.63
PG(O-36:1)	0.26	1.00	0.02	-0.44
CDP-Glycerophospholipid (38:6)	0.67	1.00	0.05	-0.39
D-Sedoheptulose 7-phosphate	0.87	1.00	0.38	-0.20
PS(41:4)	0.44	1.00	0.01	-0.49
PE(18:4(6Z,9Z,12Z,15Z)/18:4(6Z,9Z,12Z,15Z))	0.89	1.00	0.56	0.06
Dehydrotomatine	0.73	1.00	0.09	-0.38
Tin chloride (SnCl4)	0.64	1.00	0.49	-0.05
PC(14:1(9Z)/22:2(13Z,16Z))	0.28	1.00	0.02	-0.40
PC(o-18:1(11Z)/18:2(9Z,12Z))	0.62	1.00	0.14	-0.36
Arginine	0.00	0.60	0.01	-0.60
Adenosine	0.11	1.00	0.06	-0.73
6-Deoxodolichosterone	0.81	1.00	0.06	-0.53
Niclosamide	0.41	1.00	0.17	-0.35
LysoPC(20:4(8Z,11Z,14Z,17Z))	0.02	1.00	0.02	-0.79
1,2,4,8-Tetraacetoxy-7-hydroxy-3-(4-hydroxyphenyl)dibenzofuran	0.66	1.00	0.01	-1.59
3-Deoxy-D-glycero-D-galacto-2-nonulosonic acid	0.04	1.00	0.74	0.14
Glycerophosphocholine (38:6)	0.71	1.00	0.08	-0.34
N-(2-formyl-3-chlorophenyl)anthranilic acid	0.39	1.00	0.89	-0.01
5-Aminopentanoic acid	0.75	1.00	0.49	-0.10
Narciclasine	0.00	0.00	0.00	-1.00
Uric acid	0.93	1.00	0.33	-0.14
PE-NMe2(18:1(9Z)/18:1(9Z))	0.54	1.00	0.04	-0.37
PE(22:1(13Z)/15:0)	0.45	1.00	0.06	-0.37
[FA hydroxy(4:0)] 1,9S,11R,15S-tetrahydroxy-13E-prostaene	0.94	1.00	0.68	-0.24
2,3-Butanediol glucoside	0.63	1.00	0.14	-0.60
Cer(d18:0/20:0)	0.84	1.00	0.98	0.01
LysoPE(18:0/0:0)	0.20	1.00	0.01	-0.76
Camelliagenin A	0.82	1.00	0.18	-0.73
LysoPC(16:0)	0.00	0.38	0.00	-1.24
FA amino(7:0) aminoheptanoic acid	0.00	0.22	0.03	-0.44

LysoPC(20:5(5Z,8Z,11Z,14Z,17Z))	0.00	1.00	0.01	-1.02
PC(18:1(9Z)/20:1(11Z))	0.36	1.00	0.17	-0.36
Propanoylagmatine	0.96	1.00	0.68	-0.05
Lysyl-Tryptophan	0.85	1.00	0.21	-0.44
hydroxymethyl-dCDP	0.04	1.00	0.92	-0.01
Lactic acid	0.12	1.00	0.12	0.57
N-[2-(4-Chloro-phenyl)-acetyl]-N'-(4,7-dimethyl-quinazolin-2-yl)- guanidine	0.08	1.00	0.21	0.33
Deoxycytidine	0.26	1.00	0.29	0.99
Glycerophosphocholine	0.05	1.00	0.01	-0.77
1-Octen-3-yl glucoside	0.61	1.00	0.06	-0.88
Di-2-propenyl hexasulfide	0.86	1.00	0.44	0.09
PC(20:3(8Z,11Z,14Z)/20:1(11Z))	0.88	1.00	0.59	-0.21
PE(20:2(11Z,14Z)/15:0)	0.36	1.00	0.04	-0.66
N3-fumaramoyl-L-2,3-diaminopropanoate	0.13	1.00	0.16	0.83
Leucyl-Proline	0.83	1.00	0.18	-0.41
(2Z,4'Z)-2-(5-Methylthio-4-penten-2-ynylidene)-1,6-dioxaspiro[4.4]non-3-ene	0.00	0.94	0.10	0.30
Lenticin	1.00	1.00	0.68	-0.18
Tetrahexosylceramide (d18:1/16:0)	0.54	1.00	0.33	-0.21
N-(2-hydroxyhexacosanoyl)-eicosasphinganine-1-O-[D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.58	1.00	0.86	-0.03
PC(18:2(9Z,12Z)/18:2(9Z,12Z))	0.78	1.00	0.10	-0.29
Acetic acid	0.15	1.00	0.62	0.08
[ST (20:4)] cholest-5-en-3beta-yl (15S-hydroperoxy-5Z,8Z,12E,14Z-eicosatetraenoate)	0.75	1.00	0.30	-0.14
Trihydroxy-3,6,7,4'-tetramethoxyflavone	0.01	1.00	0.35	0.14
PC(20:3(8Z,11Z,14Z)/20:2(11Z,14Z))	0.70	1.00	0.31	-0.20
PG(16:0/16:0)	0.93	1.00	0.62	0.35
Dimethylarginine	0.18	1.00	0.13	-0.28
Hordatine B glucoside	0.50	1.00	0.80	0.04
N3-(4-methoxyfumaroyl)-L-2,3-diaminopropanoate	0.12	1.00	0.14	0.46
PS(35:0)	0.47	1.00	0.04	-0.60
Glyceric acid	0.07	1.00	0.52	0.67
PE(18:3(6Z,9Z,12Z)/22:0)	0.27	1.00	0.18	-0.35
PI(16:1(9Z)/18:1(11Z))	0.40	1.00	0.13	-0.40
Glycerophosphocholine (52:4)	0.88	1.00	0.18	0.12
Psidinin C	0.70	1.00	0.84	-0.11
Methyl 3-hydroxybutyrate	0.97	1.00	0.98	0.01
Gnidicin	0.14	1.00	0.35	0.60
Isocitric acid	0.78	1.00	0.56	-0.17
6-Phosphonoglucono-D-lactone	0.03	1.00	0.06	0.80
2,3-Dimethylbenzofuran	1.00	1.00	0.92	0.08
Sorbitan stearate	0.86	1.00	0.18	-0.66
cis-(homo)2aconitate	0.00	0.05	0.12	0.14
Probenecid	0.71	1.00	0.39	-0.32

Deoxyuridine triphosphate	0.08	1.00	0.29	0.72
SM(d18:0/16:1(9Z))	0.70	1.00	0.08	-0.24
PC(22:2(13Z,16Z)/16:1(9Z))	0.59	1.00	0.33	-0.28
PE(P-18:1(9Z)/14:1(9Z))	0.78	1.00	0.42	-0.09
5-(4-Acetoxy-1-butynyl)-2,2'-bithiophene	0.52	1.00	0.50	-0.19
PE(P-18:1(9Z)/20:5(5Z,8Z,11Z,14Z,17Z))	0.34	1.00	0.02	-0.73
Pyruvic acid	0.05	1.00	0.34	0.40
Creatinine	0.92	1.00	0.43	-0.11
(2alpha,3alpha,5alpha,22R,23R)-2,3,22,23-Tetrahydroxy-25-methylergost-24(28)en-6-one	0.06	1.00	0.00	-1.42
PE-NMe(16:0/16:0)	0.81	1.00	0.52	-0.23
Pipecolic acid	0.98	1.00	0.57	0.36
PG(42:6)	0.56	1.00	0.06	-0.35
N-(2-hydroxytetracosanoyl)-4,8-sphingadienine	0.96	1.00	0.95	-0.02
Diisopropyl sulfide	0.56	1.00	0.59	-0.19
Lumequoylacetone	0.82	1.00	0.12	-0.86
N-(2-hydroxyhexacosanoyl)-4R-hydroxyeicosasphinganine-1-O-[D-mannopyranosyl-alpha1-2-myo-inositol-1-phosphate]	0.71	1.00	0.41	-0.06
PE(18:3(6Z,9Z,12Z)/20:0)	0.30	1.00	0.05	-0.81
Prenol 30,32-dihydroxy-2b-methyl-bishomohopane	0.71	1.00	0.26	-0.68
2-Hydroxyundecanoate	0.99	1.00	0.97	0.03
PC(18:3(6Z,9Z,12Z)/20:3(8Z,11Z,14Z))	0.92	1.00	0.17	-0.27
4'-Methoxy-2',3,7-trihydroxyisoflavanone	0.37	1.00	0.52	-0.08
Pyrrolidine	0.14	1.00	0.02	-0.35
Cer(d18:1/24:1(15Z))	0.90	1.00	0.66	-0.25
1,6-Naphthalenedisulfonic acid	0.09	1.00	0.26	0.28
TG(41:0)	0.86	1.00	0.31	-0.20
CDP-Glycerophospholipid (38:5)	0.12	1.00	0.01	-0.61
Ganglioside GM3 (d18:1/24:1(15Z))	0.13	1.00	0.02	-0.64
SM(d18:0/16:0)	0.99	1.00	0.61	-0.20
Luteolin 7,4'-diglucuronide-3'-glucoside	0.37	1.00	0.21	-0.77
5,8-Dihydro-6-(4-methyl-3-pentenyl)-1,2,3,4-tetrathiocin	0.00	0.00	0.08	0.18
4-Hydroxybutyric acid	0.40	1.00	0.91	0.09
PS(P-39:0)	0.25	1.00	0.03	-0.51
7-methylthioheptanaldoxime	0.38	1.00	0.38	0.64
PC(18:2(9Z,12Z)/20:2(11Z,14Z))	0.86	1.00	0.36	-0.18
PI(18:2(9Z,12Z)/20:1(11Z))	0.57	1.00	0.07	-0.71
PC(18:4(6Z,9Z,12Z,15Z)/20:4(8Z,11Z,14Z,17Z))	0.89	1.00	0.79	-0.12
UDP-N-acetylmuramoyl-L-Ala-D-gamma-Glu-6-carboxy-L-Lys-D-Ala	0.92	1.00	0.25	0.08
PI (36:0)	0.74	1.00	0.27	-0.35
Stenopalustroside A	0.48	1.00	0.03	-1.35
PC(20:3(8Z,11Z,14Z)/18:2(9Z,12Z))	0.81	1.00	0.12	-0.25
6-Hydroxyluteolin 6-xyloside	0.44	1.00	0.07	-0.92

Triradylglycerol (61:8)	0.90	1.00	0.58	0.06
TG(15:0/20:0/20:0)	0.70	1.00	0.32	0.19
Piperidine	0.02	1.00	0.01	-0.42
Phosphocreatine	0.72	1.00	0.23	-0.43
5-Hexyltetrahydro-2-furanoctanoic acid	0.94	1.00	0.29	-0.34
Proline betaine	1.00	1.00	0.93	-0.07
Iodixanol	0.51	1.00	0.28	-0.15
Triradylglycerol (65:6)	0.53	1.00	0.44	-0.08
3'-O-Methyladenosine	0.36	1.00	0.44	-0.11
Kanzonol R	0.84	1.00	0.80	-0.17
5-(3-Buten-1-ynyl)-2,2'-bithienyl	0.87	1.00	0.88	0.02
1-Pyrrolinium	0.99	1.00	0.87	-0.08
SM(d18:1/18:1(11Z))	0.38	1.00	0.87	0.05
SM(d33:1)	0.72	1.00	0.14	-0.41
beta-D-Glucosyl-1,4-N-acetyl-D-glucosaminyldiphosphoundecaprenol	0.95	1.00	0.65	0.04
2,2,4-Trimethyl-3-(4-fluorophenyl)-2H-1-benzopyran-7-ol acetate	0.84	1.00	1.00	0.00
Imidazoleacetic acid riboside	0.02	1.00	0.17	0.25

Acknowledgment

This dissertation could not have been completed without the great love and support that I have received from so many people over the years.

I wish to offer my most heartfelt thanks to the following people:

To my supervisor **Prof. Dr. Dieter Jahn**. Thank you for allowing me to pursue research on topics for which I am truly passionate. You made sure to check in on me so that I stayed on the right path. I can honestly say that without your intervention this dissertation would not exist. Thank you for believing in me.

To my advisors, **Dr. Rebekka Biedendieck** and **Dr. Martina Jahn**. Thank you for the advice, support, structure and organization. You kept your office doors open for me and took control over the wheel when it was needed. You have been an ever present encouragement to me.

To my boss and advisor **Dr. Nicolas Schauer**. For all the doors you opened for me and the opportunities you have given me. We first met four years ago when I did an internship in Metabolomic Discoveries, never even thought of doing a PhD back then. Thank you for being the one of the pillars of my support system and a reason I came this far.

To my lab colleagues: **Toni**, **Stefan**, **Kathrin**, **Violetta** and **Vera**. Thank you for the community and all the conversations we shared. Laughter and sometimes tears. You all made my stay in Braunschweig so comfortable and enjoyable that the time we spent together flew by so quickly. Continue doing the great work at TU Braunschweig and wherever you go in life.

To my colleagues **Sarah** and **Lisa**. Not only are you smart, passionate and hardworking young ladies but you are funny as hell. We shared many nice moments together and I look forward to the day that the cocktails are a toast for all of us having finished our PhD work.

To my good friend and collaborator **Dr. Louisa Roselius**. You brought so much joy into my life. Not only through 9gag memes. You are the funniest and smartest mathematician I know and your presence has helped me in so many ways.

To **Metabolomic Discoveries GmbH team:** Judith, Manu, Marleen, Monique, Daniel, Josephine, Sasha and Olivia. It's been a pleasure working with you guys and learning from you.

To **Dr. Ida Retter**. You bring a new meaning to the word multitasking. I've learned a lot about planning and organization from you and have enjoyed the coffees we drank together.

To my colleagues **Maike**, **Janis**, **Matze**, **Anja**, **Johannes**, **Antonia**, **Tobi**, **Simone**, **Hilger**, **Hannes**, **Can**, **Richard**, **David**, **Kim**, **Kristin**, **Jana**, **Elia**, **Maria** and **Julia**. Thank you for the community and stimulating working environment here at Mibi. I looked forward to coming to work every day.

To my advisors **Dr. Oliver Fiehn** and **Dr. Tobias Kind**. Five months have never past by so quickly as they did this summer. Thank you for coordinating the project we worked on and all the chats we had together.

To office **colleagues at UC Davis**: Sajjan, Gert and Diego. Thank you for all the coffees and walks we had together. I look forward to coming back and creating new memories with you guys.

To my diploma thesis mentor, **Dr. Iva Rezić**. It's been few years since I was your student but the kindness and support you showed me then and continue to do stays with me still. You are such a smart and passionate researcher that I inspire to be one day.

To my friends **Juliana**, **Mojgan** and **Kate**. We started as neighbors and ended up as a family. Thank you for all the time we spent together, love and support.

To my high school friends **Sunčica**, **Jelena** and **Boris**. Thank you for all the laughs we shared together and for being my support system ever since we left high school.

To my brother **Marko**. You were my biggest fan since we started elementary school and were always supportive of almost everything I did. Thank you for all the love, laughter and tears we shared over the years.

To one of my best friends, **Dragan Janković**. Simply put, you changed my life. Almost ten years ago you became my running trainer but soon enough we became true friends. Support and love I received from you in every aspect of my life is something I am very grateful for. Thank you for all the races we partnered-up for and all the beers we drank afterwards. I look forward to the day I am back in Zagreb and we can hit a replay button.

To **my family** Antonija, Aco, Ivanka, Tin, Nikolina, Berta and Pejo. Thank you for making me part of family, all your love and support.

To my parents in law, **Zlatko** and **Asja**. Thank you for the encouragement, love and support you have shown me in the last ten years. Thank you for welcoming me into the family and all the nice moments we shared and continue to share together.

To my grandparents **Ivka** and **Frane**. You are long gone but the memory of you stays alive with me forever. Thanks to you my childhood was one of the happiest times of my life. I know you would be very proud of me.

To my dad **Simeon**. I bet you regret pushing me to become a hair dresser instead, right? Thank you for all the joy and laughter throughout my entire life. For all the car rides and lectures you have given me. Your sense of humor is what kept me going. I love you.

To my mother **Jadranka**. You have encouraged my academic interests from day one, even when I had no idea where it was leading to. Thank you for your love and measureless support, you are wonderful and most dedicated mother. I love you.

Finally, I would like to thank to the most important person in my life, my best friend and husband, **Hrvoje Torbašinović**. We came a long way together. Three countries, four cities…and that's in the last four years alone. You've been by my side for more than a decade with nothing but unconditional love and support for me. You are the smartest, most compassionate and funniest person I know and this thesis would surely not be written if it weren't for you. Thank you for changing my life for the better. I love you.

Curriculum vitae

Ivana Blaženović

Winterfeldtstraße 89

10777 Berlin

01632504518

ivana.blazenovic@gmail.com

Date of birth: 30.09.1985.

Place of birth: Zagreb

Nationality: Croatian

Work experience

March - September 2016	Visiting scholar, University of California Davis (Fiehn group), USA
May 2013 – October 2016	PhD candidate at Institute of Microbiology, TU Braunschweig, Germany
November 2012 – October 2016	Research associate at Metabolomic Discoveries GmbH, Potsdam, Germany

May 2012 – October 2012	Chemical analyst at PLIVA Hrvatska d.o.o., Savski Marof, Croatia
June – July 2009	Industrial traineeship at PLIVA Hrvatska d.o.o., Zagreb, Croatia

Education

October 2004 – March 2012	**Studies of Biotechnology, M.Sc., University of Zagreb, Croatia**
September 2000 – June 2004	XVI. Language Gymnasium, Zagreb, Croatia
June – July 2000	Summer school of English language, University of Cambridge, UK
September 1992 – June 2000	Elementary schools OŠ Savski Gaj and Bana Jelačića, Zagreb, Croatia

Languages: Croatian - native speaker

English - fluent in speaking and writing

German - advanced B2 level

Publication title:	Optimization of corrosion process of stainless steel coating during cleaning in steel brewery tanks
Authors:	T. Rezić, I. Rezić, I. Blaženović, B. Šantek
Journal:	Materials and Corrosion (2012)
DOI:	10.1002/maco.201206685